A Deep Neural Network Approach for Intelligent Crop Selection and Yield Prediction Based on 46 Parameters for Agricultural Zone-28 in Bangladesh

SUBMISSION DATE: 16.08.2018

SUBMITTED BY:

Tanhim Islam (14301112)

TanjirAlamChisty (14301096)

Prova Roy (14301137)

Department of Computer Science and Engineering

Supervisor:

AmitabhaChakrabarty, Ph.D

Assistant Professor

Department of Computer Science and Engineering

Declaration

We, hereby declare that this thesis is based on results we have found ourselves. Materials of work from researchers conducted by others are mentioned in references.

Signature of Supervisor

Signature of Authors

AmitabhaChakrabarty, Ph.D
Assistant Professor
Department of Computer Science and Engineering
BRAC University

Tanhim Islam

Tanjir Alam Chisty

Prova Roy

ABSTRACT

Agriculture is the essential ingredients to mankind which is a major source of livelihood and that provides the wide-reaching scope of working opportunities for rural people in underdeveloped or developing countries. Agriculture work in Bangladesh mostly done with old ways which directly affects our economy. In addition to, institutions of agriculture are working with manual data which cannot provide a proper solution for crop selection and yield prediction. The contribution of our thesis is to achieve the best crop selection and yield prediction in minimum cost and effort. Artificial Neural Network considered as robust tools for modeling and prediction. This algorithm aims to get better output and prediction. As well as, support vector machine, Logistic Regression, and random forest algorithm are also being considered in this thesis for comparing the accuracy and error rate. Moreover, all of these algorithms used here just to see how well they performed for a dataset which is over 0.3 million. We have collected 46 features such as – maximum and minimum temperature, average rainfall, types of land, types of chemical fertilizer, types of soil, soil moisture,soil moisture, soil consistency, soil reaction and soil texture and created our dataset for applying into this prediction process. The dataset we have considered are from past ten years (2008-2017) of Bangladesh. Therefore, based on this parameter we will predict the best possible crop selection and yield prediction intelligently.

Index term: Crop Yield Prediction, Deep Neural Network, Agriculture, Intelligent Crop Selection,

ACKNOWLEDGEMENT

First and foremost, we would like to thank the Almighty God for his never ending elegance and mercy during what ended up being one of the toughest times of our life. We would like to convey our heartiest gratitude, Dr. AmitabahaChakrabarty, Assistant Professor of the School of Computer Science and Engineering of BRAC University to give his full effort, patience, inspiration, motivation and complete guidance endlessly to accomplish our thesis successfully. It was always being a pleasure for us to work persistently with Dr. AmitabhaChakrabarty Assistant Professor of the School of Computer Science and Engineering of BRAC University with his continuous supervision and companion and providing so much support all the time to make our dream into existence. We could not have imagined a better advisor and mentor for our undergraduate thesis. Our sincere thanks go to Bangladesh Bureau of Statistics (BBS), Soil Resources Development Institute (SRDI), Bangladesh Agricultural Research Council(BARC), Bangladesh Meteorological Department (BMD) and Soil Moisture Active Passive Satellite (SMAP) for contributing us with the important data and their transparency help to conduct our thesis. The journey would not be possible without the true support of our family. We are especially grateful to our parents who supported us emotionally and financially. We always knew that they believed in us and wanted to be the best in the real world. We are grateful to them for all the support and encouragement to pursue our following dream. Finally, from our bottom of the heart, we like to give thanks to all our faculty members of Computer Science and Engineering department for supporting us throughout our undergraduate lifetime for which it is being possible to complete our thesis.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1
Introduction

In introduction, we are going to break down our entire system how it is working and a complete glimpse the functionality and features.

1.1 Introduction

Agriculture is the most important constituent for everyone around the world. For our country, it is the most employment sector [2]. Majority of Bangladeshis actually earn their livelihood by cultivating rice, jute, wheat, and other primary crops. Farmers cultivate the food and fiber we use in our everyday life. However, the process of cultivation or irrigation is doing right now by the farmers is still lag behind and have not met up today’s epoch. So our team has come up with an idea to change the trend and helping out the farmers to select the crop efficiently and maximize crop yield with minimal cost. In our thesis, we are presenting a machine-learning algorithm for agricultural zone 28 consisting of Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymensingh and Narsingdi where we are going to analyze the most effective crop selection and yield prediction for every specific region of agriculture. We are considering 6 essential major crops of Bangladesh which are Aus rice, Aman rice, Boro rice, wheat, Potato, and Jute. For the maximum agriculture prediction, we have collected about 46 major features including around 0.3 million data for our machine learning model to give our farmers the maximum cost-effective farming solution. We have used the artificial neural network to develop diverse crop yield prediction [1]. Under this neural network, we used the deep neural network with multiple hidden layers which by so far has given the best accuracy result. Furthermore, we have used Support vector machine algorithm, logistic regression and Random forest algorithm to compare with other algorithms. Environmental parameters such as average rainfall, maximum temperature, minimum temperature, humidity are exceedingly impacted to our data set [3]. We have also considered four important fertilizers, which are minerals these are Urea, Triple superphosphate, Diammonium phosphate, and MP. Moreover, we have categorized the land type (Inundation-land highland, Inundation-land medium highland, Inundation-land medium low land, Inundation-land low land, Inundation-land very low land) based on their elevation above the sea level. Furthermore, soil structure is also an important factor, which refers to the shape of soil structural unit. Based on this structure we may get to know in which land which crops yield most. The soil type is a very dominant factor for cultivating and efficiently produces a crop [2]. We have considered 19 essential soil type (Calcareous Alluvium, Non-calcareous Alluvium, Acid Basin Clay, Calcareous brown floodplain soil, Calcareous grey floodplain soil, Calcareous dark grey floodplain soil, Non-calcareous grey floodplain soil, Non-calcareous dark grey floodplain soil, Peat, Made-land, Non-calcareous brown floodplain soil, Shallow red-brown terrace soil, deep red-brown terrace soil, Brown mottled terrace soil, Shallow grey terrace soil, deep grey terrace soil, grey valley soil, Brown hill soil, grey valley soil) for our research. The ideal soil should be considered to be a loam, which means, that particular soil has an equal proportion of sand, silt, and clay. Therefore, based on these we have considered the soil moisture, texture, consistency and its reaction for better visibility of individual soil types. Finally, in our research, we tried to develop the yield prediction model using Deep Neural Network with these datasets for upmost crop yield with the highest efficiency.

1.2 Motivation

Bangladesh is predominantly an agrarian country. Due to its very fertile land and favorable weather, varieties of crop grow abundantly in this country. Agriculture sector contributes about 17 percent to the country`s Gross Domestic Product (GDP) and employs more than 45 percent of total labor force. We live in a country where we have more than 70 percent of agricultural land [5]. Our farmers are working so hard to add value in our economic growth for which they need to struggle from the early morning with a lot of stress on their head just to make our country a better world [4]. One thing actually hits our mind that if all the farmers who are actually pushing our economic growth rate and sweating for our country's better future if they could get any modern technological facilities then our country would have lead ahead. From this, our motivation arises to apply artificial intelligence in the farming sector. Our central objective was to help the farmers to cultivate land in a promoting way rather than prolonged manual procedure. In our country, about three-fifths are engaged in the agricultural sector [4]. So what we have come up with is to introduce machine learning into this sector where the machine will be going to predict based on the previous data that what need to be cultivated in this time and what shouldn't. We gathered about 46 features, which include 0.3 million databases on 2008 to 2017 yearend fitted to our machine-learning model to predict the best crop selection in a particular area where the production of agriculture will be going to boost up with minimal effort for our farmers. And another thing that also motivated us is if we started to work in this sector then our future generation will also get concern to work in this section for a sustainable agricultural development.to be cultivate in this time and what shouldn't. We gathered about 46 features which includes 0.3 million database on 2008 to 2017 yearend fitted to our machine learning model to predict the best crop selection in a particular area where the production of agriculture will going to boost up with minimal effort for our farmers. In addition, another thing that also motivated us is if we started to work in this sector then our future generation will also get concern to work in this section for a sustainable agricultural development.

1.3 Problem Statement

Data collection is always being one of the critical and inflexible blocks for doing any sort of prediction analysis. Although it appears that, it is the simplest part but in reality, it proves to be the exact opposite. At the very beginning before collecting the data, we did not have much more idea about agricultural methodology. Therefore, our first target was to gather as much knowledge as much as possible from the online source and expert people. Next, we understand about the soil structure nutrients and how crops actually yield on the field. We talked with farmers which methods they use to produce a crop. Although we did not have much more time to investigate because there are other courses which we also need to run parallel along with our thesis. Next, we surveyed Soil Resources Development Institute (SRDI), Bangladesh Agricultural Research Council (BARC) to collect data and that is what we found our first barrier because of all the data of soil nutrients, which they provided whereon, printed hard copy. Eventually making these data digitalize was our very first obstacle. Next, we need to study a lot about machine learning algorithms, how these actually work as we did not have prior knowledge about these and which model should be the best to fit our data to obtain maximum accuracy. Before using those algorithms we need to make sure our data is clean, enough before moving further which was actually a difficult task. We also had met with the SRDI and BARC scientists so many times just to make sure we are moving on to the right track. Another hindrance was to make the dataset in the right format as we have the enormous amount of dataset. We need to fix a lot of bugs and errors in our code just to train the data perfectly. We tried our best during our limited amount of time to make a predictive model, which eventually will go to help the farmers to maximize the crop production level with minimum effort and time.

1.4 Solutions

At first, we clean the unwanted observation from the dataset what we have acquired so far. There might have duplicate data because maximum amount of data was on hard copy so it is very obvious that inserting one data twice may possible. There are also some irrelevant observations which actually don't fit with the specific problem that we are trying to solve. So we cleaned it up all of those. Next we used artificial neural network to give the best prediction of our data set. We also have used Support vector machine algorithm, linear regression and Random forest algorithm. Then after some analysis we found that as there are massive amount of dataset artificial neural network gave us the better accuracy over the rest of the algorithms

1.5 Goals

Our main target is to help our country's farmer for better crop yielding with least effort. Because [4] more than three-fifth our of our country's people directly involved in the agricultural side. Therefore, this is the only strongest part of our country to easily boost up our economic growth rate. We want our farmers to get introduced to the digital world. For that, we tried to give a better prediction model by fitting in over 0.3 million previous data to calculate a better result for future crop yielding. Another goal is our farmers still do not have any sort of knowledge about soil types and nutrients. Therefore, what might be happened that cultivating the same amount of crops years after years may eventually conclude the specific nutrients to cultivate that particular crop under the soil. So if we have those data on how much nutrients does this particular soil have right now and we implement machine learning algorithm it will go to be useful for our further production level of crops. Moreover, our another goal is to meet our agricultural process to the outside world so that everyone gets inspired and start to use this proposition to the other developing countries. Finally, this entire step maybe becomes only be possible if our government actually come forward and ensure the balance of both productivity and higher returns to the farmer.

1.6 Thesis Contribution

We have inputted over 0.3 million data under 46 factors (including meteorology, varieties of chemical fertilizer, categories of land, types of soil, soil moisture, soil consistency, soil texture and soil reaction) which are the noble contribution in Agricultural research. To obtain less percentage error, we have done raw data processing, data cleaning and data normalization which drives to the maximum efficiency of our prediction and thus it will be going to help our farmers to boost up the crop yield with minimum effort. As well as, we have applied the deep neural network approach which eventually introduces a new performance benchmark of our entire system. We have perceived a highly promising result with an error percentage of less than 10%. This amount of high accuracy will finally benefit communities to make a better agricultural decision in future days. This will eventually important the farmers, government, agricultural stakeholders, policymaker and the society in our country for monitoring food security and for outlining crop yield and business.

1.7 Methodology

In our methodology part, we are going to discuss about how all the works actually being done. Firstly, we divided our dataset into two parts training and testing. Next we have clean them up before putting them to the training and testing section We have kept 80 percent of our entire dataset for training and rest of the 20 percent of our dataset for testing purpose. After that we have used deep neural network which a part of artificial neural network having many hidden layers. We got some good accuracy which is more than 95 percent. Furthermore we have used Support vector machine algorithm, linear regression and Random forest algorithm but these algorithm doesn't actually meet up the perfect accuracy that we wanted thus were giving poor accuracy result. We also have calculated variable importance by which its shows, which features, are actually dominating for proper training model. Thus, our farmers may get a solid understating about in which particular soil what type of crop they should yield.

1.8 Thesis Outline

Chapter 1 explore about the ultimate goal and what we have obtained so far.

Chapter 2 analyzes the literature review and the background study of our entire thesis.

Chapter 3 gives us a brief idea about what algorithm should be considered and choosing the perfect one.

Chapter 4 discusses the relative study and investigates the output so far we have got.

Chapter 5 Finish of the thesis and speaks about the future work plan

CHAPTER 2

Literature Review

In this chapter, we are going to discuss briefly about machine learning and what type of machine learning algorithms we have used for our research work.

2.1 Machine Learning

The world is filled with data. Machine learning has come to derive the meaning for all of that data. Moreover, it is a tool and technology that anyone can utilize to answer question with that data. In a nutshell, Machine learning is a system that provides to learn automatically and improve itself from previous confronted rather than any human involvement or assistance and being explicitly programmed.

2.2 Supervised machine Learning

Within the field of machine learning there are two different types of venture supervised learning and unsupervised learning. Now in machine learning the task is to learn a function that maps an input to an output based on the training dataset is called supervised machine learning. Most commonly used the supervised learning includes logistic regression, classification, support vector machine, artificial neural network and random forests. For making any sort of prediction analysis we at first need to create a dataset. In any sort of dataset, we categorized it by the set of features and label. Feature is actually the input and label is the output. These features can be categorical, binary or continuous [8]. We divide the dataset into two parts training and testing. Here we need to train the dataset based on the features and label that we have created and therefore predict how much accuracy we have achieved so far which we compare with the test dataset. The network model, which we created, generalizes between similar input and output patterns. As long as the model does not achieve the maximum level of accuracy, it continues its training process.

2.3 Artificial Neural Network

Human brain contains more than 100 billion micro cells, which are called neurons. These neurons carrying millions of information towards the cell body [11]. Each neuron is connected with other neurons by synapses, which carry the information [12]. In the world of computer science, artificial neural networks are inspired by human's central nerve system. This is how the comparison between computers and brains begins. The word network in the term of ANN refers to the relationship between different layers of neuron.

The artificial neural network actually works in 3 kinds of layer which are input layer hidden layer and output layer. At the very beginning before pushing any sort of data to the input layer we must declared how many hidden layers we must considered and each of the hidden layer how many neurons we should take[12]. The best way to do this is to consider neurons is to take in between of input and output layers neuron. Thus, there will be less chance that our data set is over fitted. The general equation of artificial neural network is

$$A = \sigma\left(\sum w * a + b\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(i)}$$

In this equation:

- A = Activation function refers how active that neuron actually is.
- σ = sigmoid is used to add nonlinearity of the model.
- w = the weighted sum of previous neurons
- a = all the input activation of previous layer
- b = bias which allows to shift the activation function left or right to fit the prediction with better data

$$A = \frac{1}{1+e^{-x}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(ii)}$$

Now to initialize our neural network we at first fit the data to the very first input layer, which randomly creates a bunch of weights from one neuron to the next. There is a bias which also randomly generated for shifting our function and get better accuracy. Every time when we move to the next node, it multiplies the weight of its previous synapse along with the activation node and adds the bias. Next, we put everything in a sigmoid Function [13]. The benefit of sigmoid function is, it is nonlinear and has a smooth gradient too. Therefore, in sigmoid function if we change a value in X then it has a tendency to bring the Y value to the end of the curve, which is actually great.

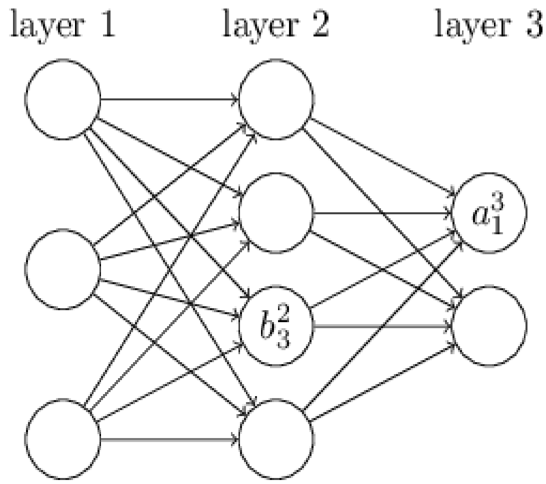

Figure 2.3.1: A simple Artificial Neural Network model [12]

We next calculate the actual cost. Now what cost refers is that it actually shows how close we have gone to our actual output result. The lower the cost rate is the better accuracy we will get. The cost is calculated by-

Cost = $(\text{Output achieved} - \text{actual output})^2$… (iii)

After calculating the total cost, we need to improve our model where we will use the feed forward back propagation algorithm. This algorithm back propagates to change the value of weights and biases to minimize the cost function. Back propagation works in several ways to improve the activation of the particular output neuron. These are –

- Increase the bias
- Increase the weights
- Increase the activation of previous layer

By following these steps we can easily minimize the cost and thus get better accuracy in the output layer. Here is a basic work for calculating the errors with the perspective of weight [10].

$\frac{\partial E}{\partial wi}=\frac{\partial E}{\partial y} * \frac{\partial y}{\partial net} * \frac{\partial net}{\partial wi}$

$\frac{\partial E}{\partial wi}$= Changes in error when the weights are changed

$\frac{\partial E}{\partial wi}$= Changes in error when the output are changed

$\frac{\partial E}{\partial wi}$= Changes in output when the weighted sum are changed

$\frac{\partial E}{\partial wi}$= Changes in weighted when the weights are changed

Deep Neural Network

In our thesis, we have implemented Back propagation method to train our deep neural network where we have used 3 hidden layers to calculate the total cost of our output. Sometimes handling enormous amount data by using only 1 hidden layer may not give a high accuracy. Although adding more hidden layers may increase computational cost but it can better generalize to big data.

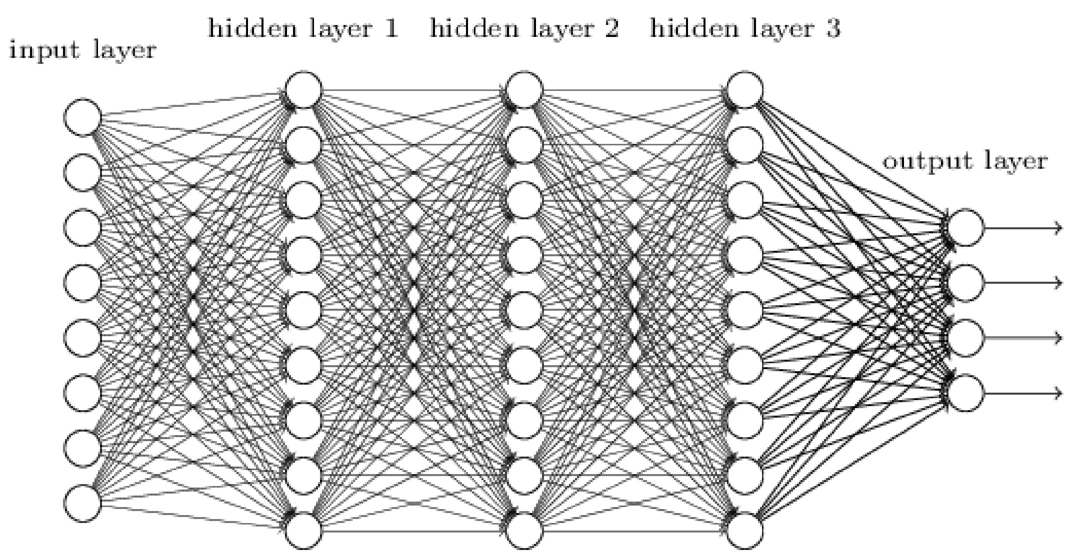

Figure 2.3.2: A simple Deep Neural Network model [17]

2.4 Pseudo code for Back propagation to train Deep Neural Network

Algorithm 1: Back Propagation Deep Neural Network Algorithm

1. Initialize network weights (often small random values)
2. **do**
3. **forEach** training example ex
4. Prediction = neural-net-output(network, ex) // *forward pass*
5. Actual =teacher-output(ex)
6. Compute error (prediction - actual) at the output units
7. Compute Δw_h for all weights from hidden layer to output layer // *back pass*
8. Compute Δw_i for all weights from input layer to hidden layer // *back pass*
9. Update network weights // *input layer not modified by error estimate*

10. **until** all examples classified correctly or another stooping criterion satisfied
11. **return** the network

2.5 Support Vector Machine

The main objective of SVM is to design a hyper plane that classifies all training vectors in two classes. Now to separate the classes we draws some hyper planes and among of those we choice the best hyper plane that leaves the maximum distance from the both classes. The equation of the hyper plane is-

$$g(\overline{X}) = w * x + w0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(iv)}$$

Now here if g(x-bar) > 1 then it refers to upper class or class 1 and if g(x-bar) < -1 then it refers to lower class or class 2 The total margin is computed by-

$$\frac{1}{||w1||} + \frac{1}{||w2||} = \frac{2}{||w||} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(v)}$$

If we can minimize the ||w|| term then we can maximize theseparability

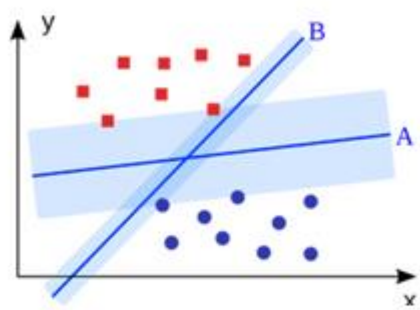

Figure 2.5.1: A simple Support vector machine model [14]

From the above figure it's easier to understand that the margin value of A is higher than B so we should consider hyper plane A for differing among the classes.

2.6 Pseudo code for Support Vector machine

Algorithm 2: Support Vector Machine

1. Initialize $y_i = Y_I$ for $i \in I$
2. **REPEAT**
3. Compute SVM solution w, b for data set with imputed labels
4. Compute outputs $f_i = \langle w, xi \rangle + b$for all x_i in positive bags
5. Set $y_i = \text{sgn}(f_i)$ for every $I \in I$, $Y_I = 1$

6. **FOR** (every positive bag B_I)
7. **IF** ($\sum_{i \in I}(1 + y_i)/2 == 0$)
8. Compute $i^* = \text{argmax}_{i \in I} f_i$
9. Set $y_{i^*} = 1$
10. **END**
11. **END**
12. **WHILE** (imputed labels have changed)
13. **OUTPUT** (w, b)

2.7 Logistic Regression

Logistic Regression [15] is a little bit similar to Linear Regression in the sense that both have the goal of estimating the values for the parameters/coefficients, so the at the end of the training of the machine learning model we got a function that best describe the relationship between the known input and the output values. The benefit of using this algorithm is it can be used to fit complex nonlinear datasets which means we can move on to more complexity boundaries by fitting complex parameters by using higher order polynomial. In logistic regression, a complex formula is required to convert back and forth from the logistic equation to the OLS-type equation. The logistic formulas are stated in terms of the probability that Y= 1, which is referred to as p .The probability that Y is 0 is 1-p.

$$ln = \left(\frac{\mathrm{p}}{1-\mathrm{p}}\right) = B0 + B1 * X \ldots\ldots\ldots\ldots\ldots \text{(vi)}$$

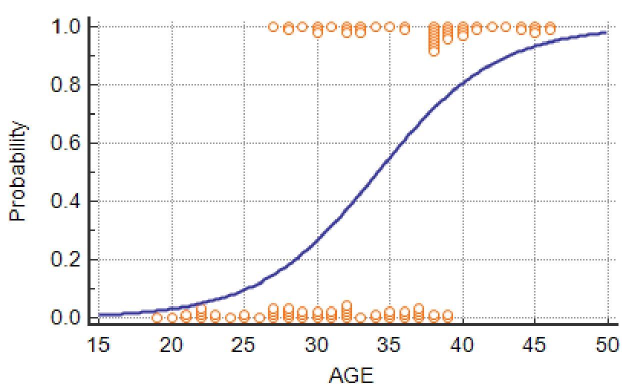

Figure 2.7.1: A simple logistic regression model [15]

2.8 Random Forest

Random forest algorithm is a one of the most popular and most powerful supervised Machine Learning algorithm in Machine Learning that is capable of performing both regression and classification tasks. As the name suggest, this algorithm creates the forest with a number of decision trees. In general, the more trees in the forest the more robust the prediction. In the same way in the random forest classifier, the higher the number of trees in the forest gives the high accuracy results. For example if 500 trees are grown and 400 of them predict that a particular pixel is forest and 100predict it is grass the predicted output for that pixel will be forest.

2.9 Pseudo code for Random Forest

Algorithm 3: Random Forest

Precondition: A training set S :=($x1$, $y1$),….. (x_n,y_n), features F, and number of trees in forest B

function RANDOMFOREST(S,F)
H ← **ϴ**
for$i$∈1,…,B do
$S^{(i)}$← A bootstrap from S
h_i← RANDOMIZEDTREELEARN($S^{(i)}, F$)
H←H∪ {h_i}
end for
return$H$
end function
function RANDOMIZEDTREELEARN(S, F)
    At each node:
f← very small subset of F
        Split in best feature in f
Return The learned tree
end function

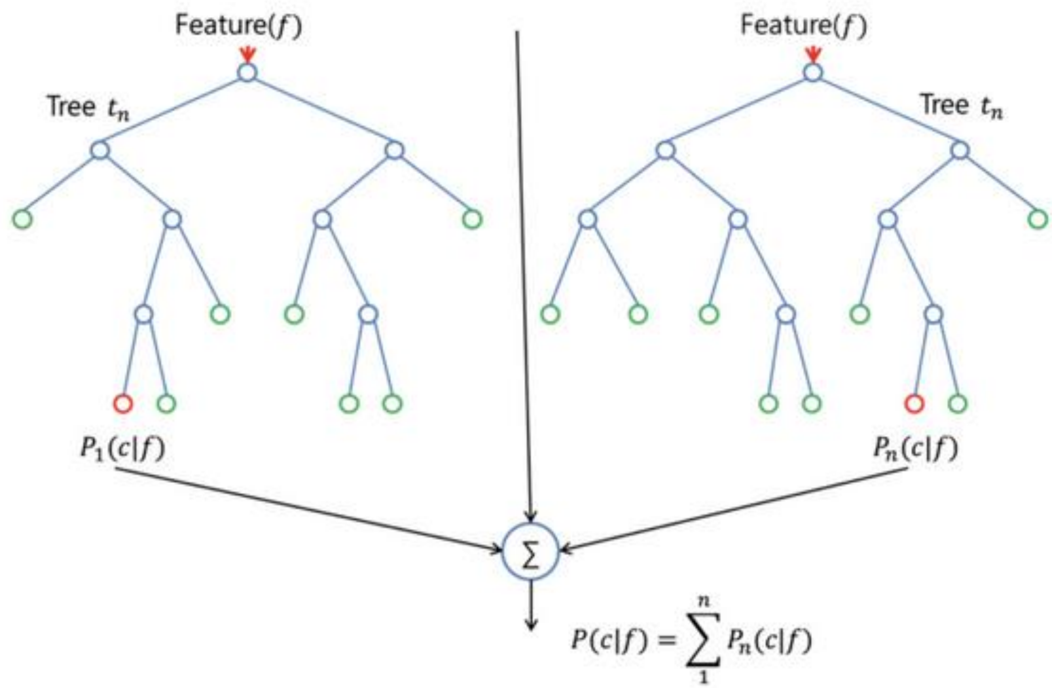

Figure 2.9.1: A simple random forest model [16]

CHAPTER 3

Data Collection and Table Formation

In this section we are going to discuss about the most strong point of our research work which is how we have collected these massive amount of data and created our own dataset.

3.1 Data Sources

Before starting our thesis one thing which always bugged us was data source. Machine learning is all about learning from dataset and forecast future trends. As we are considering agricultural zone 28 so our primary target was to know where we can get agricultural data for our research work. Online surveys can be used to understand about the area under study. Therefore, at the very beginning we searched some of the organizations who are directly working on this field. Firstly we get to know about Soil Resource Development Institute (SRDI), which works currently working on soil nutrients fact. As the organization wasn't very far from our place so we decided to move on. We firstly took appointment of Mr. Bidhan Kumar Bhandar director of SRDI and explained him about our needs and the benefit that our country will get from their data. Furthermore we also moved to Bangladesh Agricultural Research Council (BARC) where we get the chance to meet Md. KabirIkramulHaque the Chairman of BARC. Next we proposed our research and asked them to provide their dataset to us. Later on they gave us the book of their dataset collection. People who had already done their research in this sector mostly considered only the Meteorology and Agricultural output section but we thought only depending on these features are not enough to give a perfect crop yield prediction that's why our team worked hard to find other important factor which dominates the crop yield output. Next we move on to Bangladesh Bureau of Statistics-Government (BBS) and Bangladesh Meteorological Department (BMD) website where we have got other data sources. But all this dataset was not ready for implementing machine learning algorithm.

3.2 Data Collection

Regardless of the approach taken, data collection and proceeding is always being diligent and time consuming**.** So far we found out 4 different organizations who are assigned on different responsibilities. Our next job was to collect all the data from these sources. Come to the SRDI we got access of SRDI library where there was thousands of books on different agricultural zones. We collected 70 books from year 2008 – 2017. Unfortunately the books we have collected were written in Bangla digit. So the initial task for us was to make these manual data digitalize. Next up we collected the data from BARC which was scattered because the yearend book which they publish has many chapters and from each chapter there are different types of data format so we had to arrange those as well. Same thing also went on BBS and BMD dataset.

3.3 Data Processing

After getting all the data we have found out that there were about 46 different types of features which actually create a great impact for agricultural crop yield production. The data we have finally got is 206126 (roughly under 0.3 million). The first problem was giving the input of 70 books which we have taken from SRDI. This actually took 3 months to make these data digitalize. Another thing was, we scaled those data.Since, the value was in different unit so we have changed those as well. Moreover, we cleaned the raw data, because there were many null values. Thus, we format all these data in a proper way before plugin those data in our machine learning algorithm. In addition to, many columns were in categorical format. Therefore, we used 'LabelEncoder' function to convert it into a unique numerical value. Again, all these data was collected from several sources so we compiled it into a single tabularized form. To apply machine learning algorithm, we composed this tabularized excel format into comma separate value format. Then, we loaded our dataset into system and applied rescale data technique, standardize data technique and normalize data technique. As well as, we implied 'OneHotEncoding' technique to represent categorical variables as binary vectors. Finally, 'StandardScaler' function was used to do proper feature scaling in our predictive model.

3.4 Data Description:

3.4.1 Meteorology

These data was collected from Bangladesh Meteorological Department (BMD) and the most common features to work for any agricultural output. Here are the features that we have considered for our research.

- **Average Rainfall:**

  Bangladesh is a lowland area. So during heavy rainfall there might be heavy possibility of having flood. If a plants soil receives excessive amount of water during rainfall then the plant cannot get enough oxygen from soil. So in order to achieve great crop yield we need have moderate rainfall to maximize our crop productivity level.

- **Maximum Temperature**

  The temperature in Bangladesh is drastically increasing every year. Plants have a photosynthesis process which works under a certain temperature level. So we can understand this parameter is one of the key factors for predicting crop yield.

- **Minimum Temperature**

  This parameter is also very important for predicting crop yield because with the help of average minimum temperature of a region it will give us a solid understanding whether the plant is growing optimally or it is hampering the plant growth.

- **Humidity**

  Humidity is another important feature for our data analysis because it has a great impact on plant growing. Basically what humidity does is it helps the water cycle process of our environment. As plant transpire in a certain humidity level so we can get to know how much humidity is changing in our agricultural zone and how it is effecting the growth rate.

3.4.2 Types of Land

These data was collected from Soil Resource Development Institute (SRDI) which was written in Bangla digit in book. So we considered these features, because depending on the land different types of land production of crop differs a lot. Here in our country have Inundation land which is covering of a land area by water brought about by natural causes. Here are the features that we have collected for our thesis.

- **Inundation-land highland**

  This type of land is actually above normal inundation level.
- **Inundation-land medium highland**

  This land normally inundated up to about 90 cm deep
- **Inundation-land medium low land**

  This type of land inundated up to 90 to 180cm deep.
- **Inundation-land low land**

  This type of land inundated up to 180 to 300cm deep
- **Inundation-land very low land**

  This type of land inundated deeper than 300cm.
- **miscellaneous land**

  This type of land have little amount or no natural soil which is not suitable place to grow crop.

3.4.3 Chemical Fertilizer

These data was collected from Bangladesh Agricultural Research Council (BARC). Depending on the chemical fertilizer on the soil we can analyze how much impact they are actually creating when this value varies. Here are the features that we have collected for our thesis.

- **Urea**

  Urea is a white crystalline soil containing 46% nitrogen, and it can be considered an organic fertilizer. When we applied urea to the soil both will convert to the ammonia and nitrogen needed by plants.

- **Triple superphosphate**

  Triple superphosphate is one of the first high analysis phosphorous fertilizers. Fertilizers often supply significant amounts of phosphorus, along with nitrogen and potassium. Super phosphate is a common synthetic phosphorus fertilizer. It contains a large percentage of phosphate and it needs to apply carefully to ensure the plants can use it perfectly.

- **Diammonium phosphate**

  This is the most widely used chemical phosphorus fertilizer. It has very high nutrient which is excellent for farming and crop yielding.

- **MP**

  This fertilizer is also very important for growing plants. This fertilizer along with other three plays a very important role for crop yielding.

Table 1: Dataset sample up to chemical fertilizer

District	year	area	avg_rainfall	max_temperature	humidity	urea	tsp	DAP
dhaka	2008	793	2385	34.2	71	25967	8262	1573
dhaka	2009	1125	1930	35.6	66	29300	8041	1578
dhaka	2010	901	1523	35.5	55.3	27967	8675	1688
dhaka	2011	650	1776	23.4	68.2	36991	9386	1628
dhaka	2012	628	1329	30.8	58	32966	8166	1925
dhaka	2013	502	1590	30.9	63	30332	9188	1941
dhaka	2014	477	1399	31.1	64.7	35321	9233	1926
dhaka	2015	692	2166	34.2	60.2	30750	9544	1586
dhaka	2016	654	1562	27	65	36712	9094	1987
dhaka	2017	655	1640	31	69	37382	8782	1903
gazipur	2008	2608	2197	30.2	71	29264	9225	5721
gazipur	2009	2669	1912	31	66	29059	9658	5604
gazipur	2010	2487	1181	29.7	55.3	29295	9435	5967
gazipur	2011	2332	1777	31	68.2	29457	9866	5831

3.4.4 Soil Type

These data was collected from **Soil Resource Development Institute (SRDI).** Soil is one of the most important factors for making decision. This soil type will go to help us to take important decision for future prediction. These are the features that we have collected for our research work.

- **Calcareous Alluvium**

  This particular soil has raw sandy and silt alluvial deposits. In addition they are slightly moderate calcareous due to presence of calcifies derived mostly from the gangetic sources. They are actually saline in coastal areas.

- **Non-calcareous Alluvium**

  It's basically raw sandy and silt alluvial deposits, usually classify either from the surface or below the cultivated topsoil in the active floodplain areas. They are neutral to alkaline in reaction. They are saline in the coastal tidal areas. Dhaka, Gazipur, Mymensingh, Narsingdi, Tangail land has this characteristics.

- **Acid Basin Clay**

  This soil is very strongly acid, grey to dark grey heavy plastic clays. They are actually seasonally deeply flooded and have heavy consistence. Dhaka, Gazipur, Mymensingh, Narsingdi, Tangail has this sort of soil.

- **Calcareous brown floodplain soil**

  Mostly Dhaka district has this sort of soil. This soil is calcareous, brown silt loams to light silt clays. Locally they are leached of lime up to a depth of 1m from the surface.

- **Calcareous grey floodplain soil**

  In our agricultural zone this soil is not present though but this soil plays some vital role as well. It is structured, grey silt loams to silt clays, calcareous from the surface or at shallow depths.

- **Calcareous dark grey floodplain soil**

In our research from the 7 districts only Dhaka district got this type of soil presence. This soil is structured dark grey silt day loams to heavy clays occurring in basins and on low ridges. These soils are calcareous within a depth of 1.2m below the surface. Clays are highly cracking when dry, drought prone have heavy consistence. They become saline in the day season.

- **Non-calcareous grey floodplain soil**

  All of the 7 districts of our agricultural zone have the presence of this sort of soil. This soil is prismatic and blocky structured predominantly grey sandy loams to silt clay loams. They become saline in dry season

- **Non-calcareous dark grey floodplain soil**

  This soil is also being present in those 7 districts. It is structured on dark grey soils on old flood plain ridges and clay in basins. Slightly acid to somewhat alkaline in reaction. The basin clays have heavy consistence.

- **Peat**

  From agricultural zone 28 only Dhaka district has peat. It is highly organic dark colored soils developed under soil developed under submerged conditions. It drained and allowed to dry out peat shrinks irreversibly, this causing crack of the soil.

- **Made-land**

  This particular soil is a thin layer of material which is actually covering the earth's surface and is formed from the weathering of rocks.

- **Non-calcareous brown floodplain soil**

  From agricultural zone 28 only in Narsingdi and Mymensingh district contains these type of soil. This soil is non-calcareous brown sandy loams to clay loams occurring in the old Himalayan piedmont plain. These soils are slightly too strongly acid in reaction.

- **Shallow red-brown terrace soil**

The color of this soil is either red or brown or the combination of both. Usually this soil is strongly acid and structured. They mainly occur on the narrow terraces and locally. In Dhaka, Gazipur, Mymensingh and Tangail have this type of soil.

- **deep red-brown terrace soil**

  Dhaka, Gazipur, Mymensingh and Tangail and Narsingdi have this particular soil. This soil is brown to red and slightly on strongly on acid side and finely structured. It is gradually intergrading into mixed red, black and pale brown clay.

- **Shallow grey terrace soil**

  Dhaka, Gazipur, Mymensingh and Tangail contain this particular soil type. The color of this soil is kind of whitish grey and it is slightly on strongly acid.

- **deep grey terrace soil**

  This particular soil also has whitish grey, speckled with brown or red mottles and it is strongly to acid side. Dhaka, Gazipur, Mymensingh and Tangail have this particular soil.

- **grey valley soil**

  From agricultural zone 28 Mymensingh district has this particular soil type. This soil is structured, grey sandy loams to clays strongly acid. The problem with this soil is it is often affected by flash flood.

- **Brown hill soil**

  From agricultural zone 28 Only in Mymensingh district this particular soil type is available. This particular soil is brown sandy loams to clay loams and mostly it is slightly too strongly reacted with acid.

3.4.5 Soil Information

These data were collected from different yearbook of Bangladesh Bureau of Statistics-Government (BBS) and one problem was all of these data were in scattered format. So we need to make sure all of the data are in right format. These are the factors we have considered for our thesis work.

- **Soil Moisture**

  Soil moisture is a key variable in controlling the exchange of water and heat energy between the land surface and the atmosphere through evaporation and plant transpiration. Soil moisture plays an important role in the crop production or crop yield

- **Soil Consistency**

  This is basically the strength with which materials are soil materials or nutrients hold together. If the soil consistency is wet the nit can be considered as stickiness. This is a really important parameter for our prediction model.

- **Soil Reaction**

  Soil reaction or the soil pH is the same thing which is another important feature that measures the indication of acidity or alkalinity of a particular soil. The scale is between 0 to 14. pH 7 is the neutral point.

- **Soil Texture**

  Soil texture is the determination of soil classes based on their physical texture. There are actually three different types of sol particles.

  - Sand
  - Silt
  - Clay

  These are basically the soil particle size and the most important factor for training our model.

3.4.6 Agricultural Output

These data are the label data that we are going to predict actually. These data were also being collected from Bangladesh Bureau of Statistics-Government (BBS). The problem with these data was all of these data were in different unit so we need to scale it to bring all of those in one general format. Here are the label data we have considered so far.

- **Crop Yield**

  This is the main agricultural output which defines how much crop is producing in a particular land or area. This is an important parameter because just looking at the production unit we cannot come to conclusion that this crop is perfect in this particular area.
- **Area of Individual Crops**

  This output considers that how much has been taken to produce one particular type of crop.

- **Production of Individual Crops**

  This result shows us how much crop at the end actually produced. But the difference is in a certain area which crop actually yields the most. So this is an important feature for our labeling.

Table 2: Dataset sample of soil type

Calcareous Alluvium	Noncalcareous Alluvium	Acid Basin Clay	Calcareous Brown Floodplain Soil	Calcareous Grey Floodplain Soil	Calcareous Dark Grey Floodplain Soil	Peat	Made-Land
4869	6782	6253	7691	26289	28109	13900	7048
4912	6831	6284	7701	26346	28881	14217	7049
4925	6784	6267	7720	26343	28493	14118	7157
4885	6825	6282	7719	26346	28425	14277	7065
4930	6862	6295	7707	26294	28895	14148	7005
4877	6877	6254	7707	26311	28987	14205	7138
4952	6832	6297	7703	26322	28199	14095	7042
4909	6848	6281	7714	26371	28719	14214	7038
4903	6825	6258	7718	26304	28519	14088	7017
4923	6851	6254	7718	26338	28746	14155	7005
0	648	28245	0	0	0	0	0
0	760	28708	0	0	0	0	0
0	740	29734	0	0	0	0	0
0	664	28723	0	0	0	0	0

Table 3: Dataset sample up to crops

soil moisture	soil texture	soil consistency	soil reaction	aus	aman	boro	wheat	potato	jute
124237	123164	144439	114506	804	9691	233939	1129	33679	29825
124180	129365	142913	106664	1492	28872	217990	2306	42224	55858
124271	128411	144273	118789	1353	28772	203305	901	49029	63147
124170	119648	145716	108940	912	26151	226879	710	45670	67453
124357	128189	140355	109144	927	23809	204100	565	41564	58982
124206	125283	143752	110160	752	27797	190646	513	40571	57230
124158	124955	140524	112110	948	25676	200560	710	41062	52200
124203	129062	145234	116778	1413	15604	205638	716	51989	58234
124269	121505	140658	109278	1431	27419	211933	657	47167	54177
124287	122824	145954	116000	1407	27671	218997	742	41497	54590
154669	157481	154669	154669	4328	93956	208434	479	4300	15072
154625	158865	154802	158445	4199	105450	217819	447	4416	9662
154839	153959	156059	153197	4086	125245	247999	365	5475	10338
154779	153369	152296	157790	3869	125950	220675	291	5054	17556
154740	154224	151254	151765	4103	116994	250778	273	4601	6993
154900	156829	159754	149331	3730	111453	237235	238	3353	10458
154602	154227	155507	155595	6324	97502	228161	205	2711	13464
154879	154597	158324	149975	4071	101542	231718	156	3335	12378

CHAPTER 4

System Architecture

This chapter will going to discuss briefly about the system specification with the help of a diagram to give better understanding how the system is working.

4.1 Six Major Crops

According to Bangladesh Bureau of Statistics six major crops in Bangladesh are

1. **Aus Rice**: It is one of the major crops in Bangladesh, which has been contributing to agricultural production alongside with other to rice (Aman and Boro) crops [10].
2. **Aman Rice**: It is one of the main crops in Bangladesh, which is the second largest rice crop in the country in respect to the volume of production while boro tops the production. The production of Aman depends on the weather condition of the country and farmers usually cultivate Aman in their land [10].
3. **Boro Rice**: Boro is the most important and single largest crop in Bangladesh in respect of volume of production which is continuously contributing to higher rice production in the last successive year [10].
4. **Wheat**: Wheat is one of the most valuable winter crops in Bangladesh, which is temperature sensitive. Wheat is the second most important grain crops after rice [10].
5. **Jute:** Jute was found grown in Bangladesh almost solely as a rain fed crop without any irrigation or drainage provisions. Jute is one of the important sectors in Bangladesh economy [10].
6. **Potato**: Potato provides a critically important element in the diets of many people in Bangladesh [10].

4.2 Sample Area:

In our thesis, we are choosing Agricultural Zone-28 as a sample area which consisting of:

- Dhaka
- Gazipur,
- Narayanganj
- Tangail,
- Kishorgonj
- Mymenshing
- Narsingdi

Where we are going to analyze the most effective crop selection and yield prediction for every specific region of agriculture. We are considering 6 essential major crops of Bangladesh which are Aus rice, Aman rice, Boro rice, wheat, Potato and Jute. For the maximum agriculture prediction we are taking about 46 major features including around 0.3 million data for our machine learning model to give our farmers the maximum cost effective farming solution. [1] We have used artificial neural network to develop diverse crop yield prediction. Under this network we used deep neural network with multiple hidden layers which by so far has given the best accuracy result. Furthermore, we have used Support vector machine algorithm, linear regression and Random forest algorithm to compare with other algorithms

4.3 System Workflow

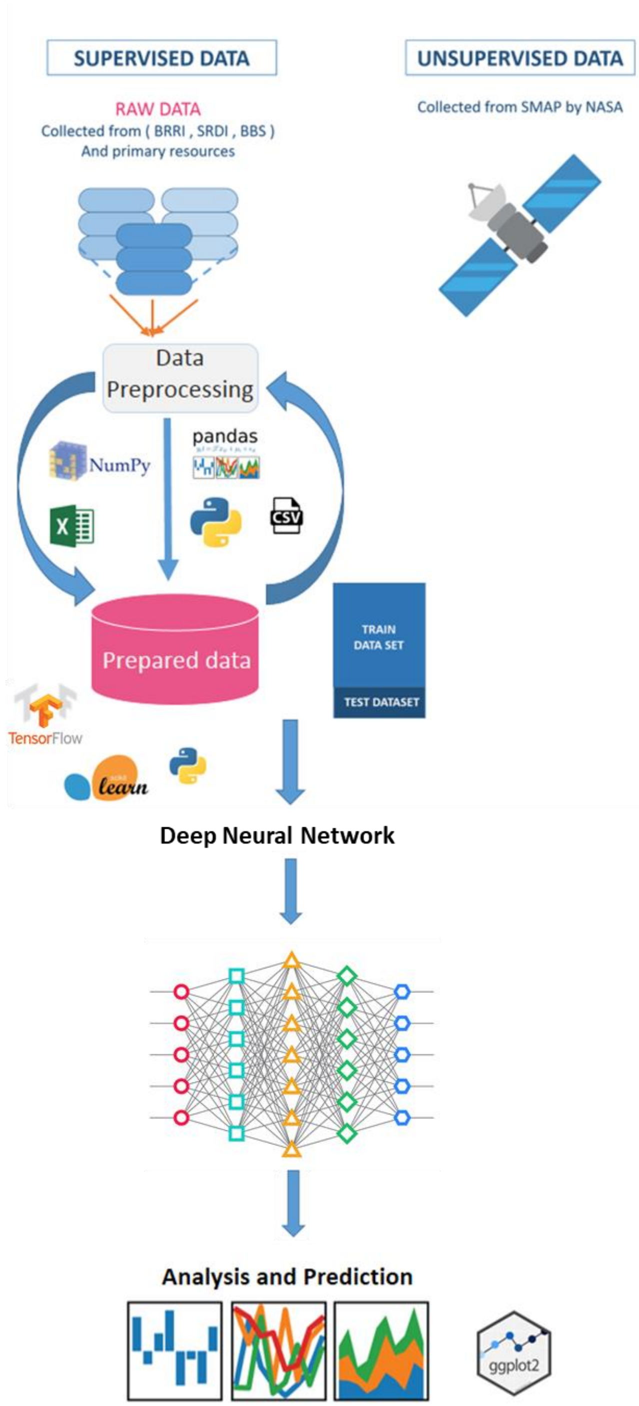

Figure 4.3.1: Workflow

4.4 Room For Improvement:

Our future work plan is to get precise agricultural weather through remote sensing which will be going to boost up the prediction of crops undoubtedly. The reason behind moving towards to remote sensing because in reality it is really back-breaking to carry out accurate surveys of the surface area of planted fields. So this is the place where remote sensing can be used to discover accurate data.

- Soil Moisture Active Passive satellite (SMAP)

[13] NASA has launched Soil Moisture Active Passive satellite (SMAP) in 31st January 2015 to measure the land surface soil moisture and to provide accurate climate data. The SMAP measurement system consists of a radiometer, which is a passive instrument, and synthetic aperture radar, which is an active instrument operating with multiple polarizations in the L-band range. The combined active and passive measurement approaches takes advantages of the spatial resolution of the radar and the sensing accuracy of the radiometer. The active and Passive sensor sense the conditions in the top 5 cm of soil through moderate vegetation cover to yield globally mapped estimates of soil moisture and its freeze-thaw state. Soil moisture observations from SMAP will lead to improvements in crop yield forecasts and will enhance the capabilities of crop water stress decision support system for agricultural productivity. For our purpose if we can integrate this unsupervised data through remote sensing we can get some extra benefit such as

- We can able to predict Flood for which thousands of crops are destroying and billions of dollars are losing in every single year.
- Drought can result crop failure so what SMAP can help us that it monitors soil moisture and provides critical information for drought early morning.
- SMAP will help us by driving data of the amount of water available to evaporate from the land surfaces, which will be going to lead us to make decision to better agricultural

CHAPTER 5

Dataset and Result Analysis

In this section, we are showing the result analysis throughout the year and have shown the comparison among the different algorithms. The comparison graphs are in the following page.

5.1 Data Analysis of Maximum Temperature

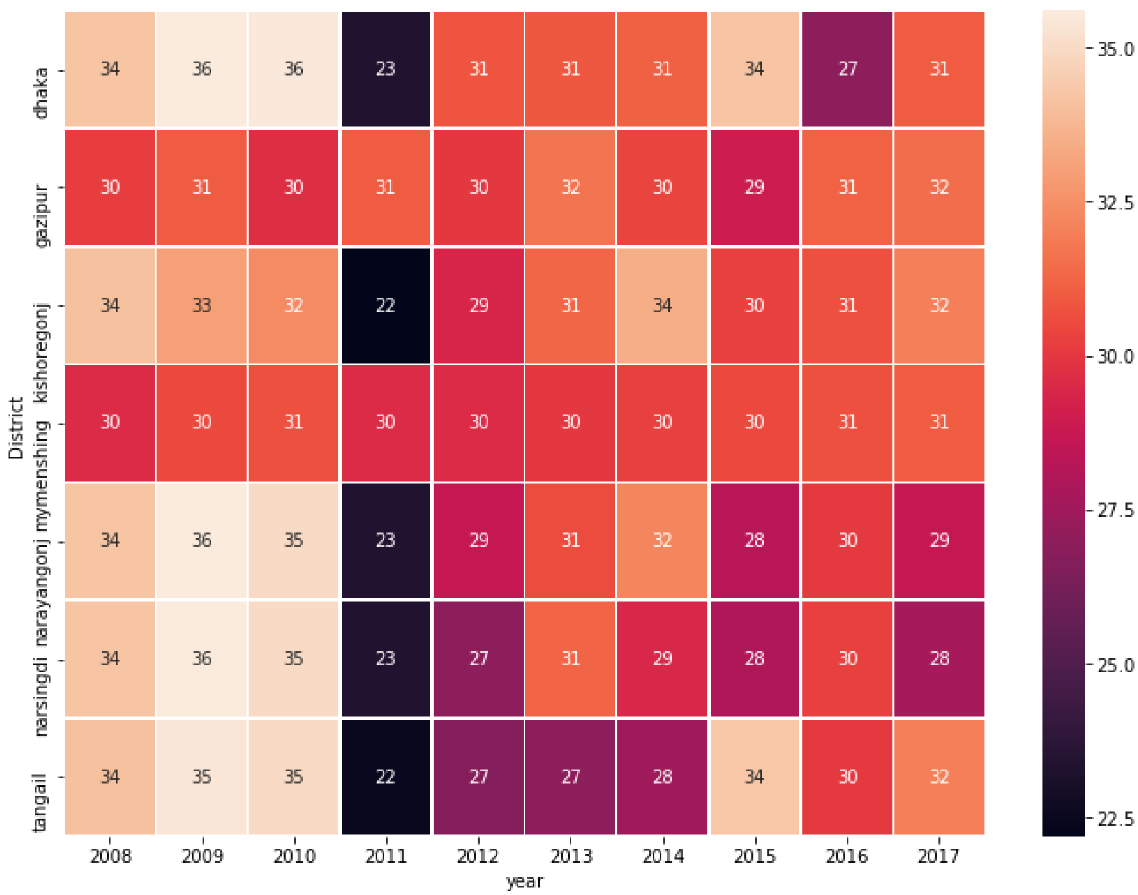

Figure 5.1: Analysis of Maximum Temperature from year (2008-2017) against district

In the figure 5.1, describes the utmost temperature in Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymenshing and Narsingdi from year 2008 to 2017. In the horizontal axis, we have set year (2008-2017) and in the vertical axis we have set district.

Annual maximum temperature was recorded from 22.5 degree Celsius to 35 degree Celsius. As we can see in Dhaka, city during 2008 it was about 34 degree Celsius and almost most of the 7 districts are in between 30 to 34 degree Celsius. But as we move forward the temperature increases 1 to 2 units in almost every district till 2010. But the temperature drastically got down during 2011 expects Gazipur and Mymenshing. Eventually the temperature again increases and being stable from year 2012 to 2017.

5.2 Data Analysis of Minimum Temperature

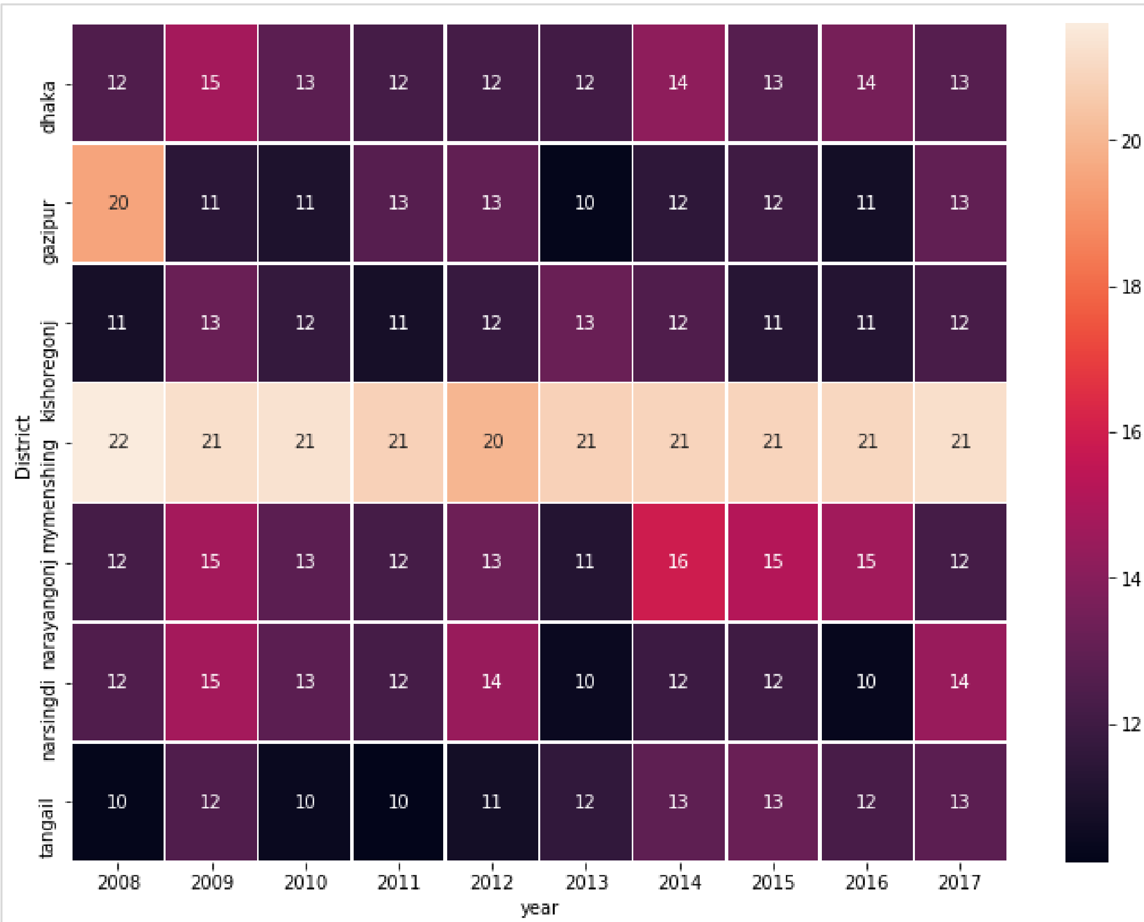

Figure 5.2: Analysis of Minimum Temperature from year (2008-2017) against district

In the figure 5.2, describes the utmost annual minimum temperature in Dhaka, Gazipur, Narayangonj, Tangail, Kishoregonj, Mymenshing and Narsingdi from the year 2008 to 2017.In the horizontal axis, we have set year (2008-2017) and in the vertical axis we have set district. Annual minimum temperature was recorded from 12 degree Celsius to 20 degree Celsius. As we can see that in Mymenshing district minimumtemperature is recorded between 20 degree Celsius to 22 degree Celsiuswhich is higher that other districts.Also, In 2009 , annual minimum temperature increases 3 unit in Dhaka, narayangonj and narsingdi district.Furthermore, in Dhaka district minimum temerature ranges from 12 degree Celsius to 15 degree Celsius, in Gazipur district minimum temperature ranges from 11 degree Celsius to 20 degree Celsius, in Kishoregonj district minimum temperature ranges from 11 degree Celsius to 13 degree Celsius, in Mymenshing district minimum temperature ranges from 20 degree Celsius to 22 degree Celsius, in Narayangonj district minimum temperature ranges from 10 degree Celsius to 15 degree Celsius and in Narsingdi district it varies from 10 degree Celsius to 13 degree Celsius.

5.3 Data Analysis of Average rainfalls

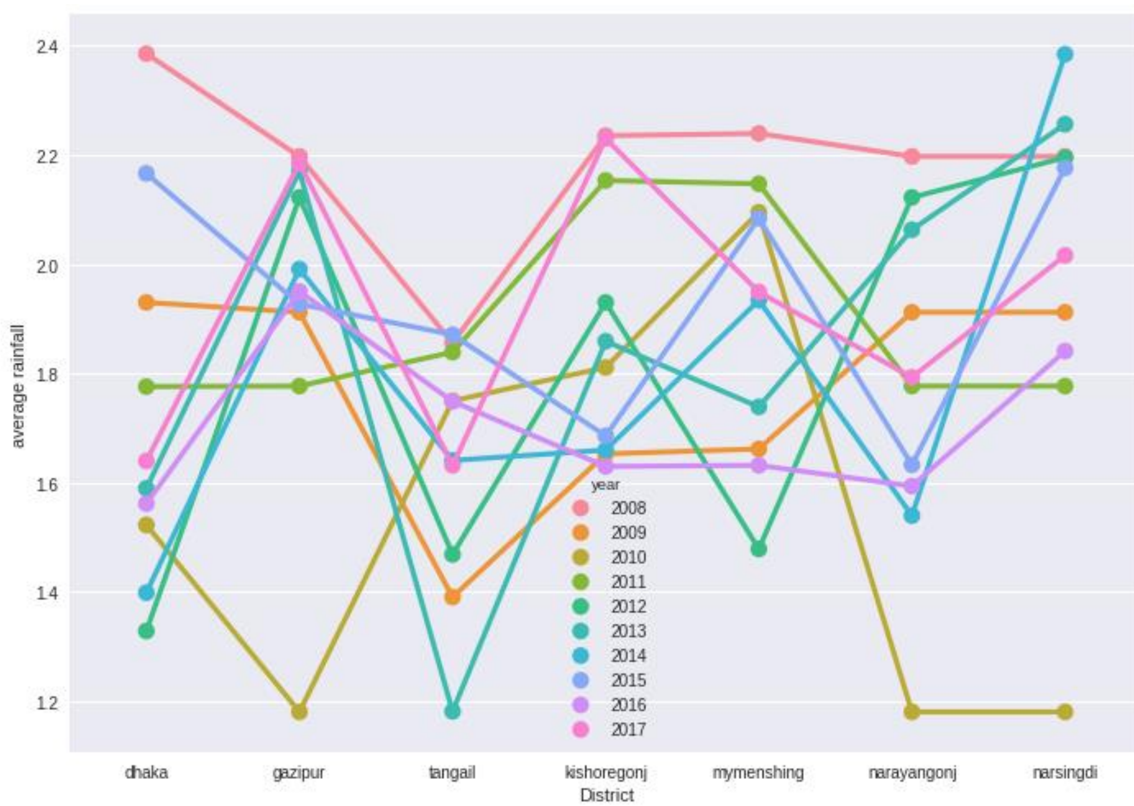

Figure 5.3: Analysis of Average Rainfall

In figure 5.3, displays the result analysis of average rainfall in Dhaka, Gazipur, Narayangonj, Tangail, Kishoregonj, Mymenshing, and Narsingdi district from the year 2008 to 2017. In the horizontal axis, we have set districts (Agricultural Zone-28) and in the vertical axis, we have set average rainfall (mm). Annual average rainfall was recorded from 12 mm to 24 mm. As we can visualize that Dhaka and Narsingdi have received the highest amount of rainfall during 2008 and 2014, which was about 24mm and performs the least in 2012 and 2014 which was about in between 12mm and 14mm. The most static rainfall happened in Kishoregonj district where the average rainfall was about 16mm to 22.3 mm. But in 2010 almost all of the districts got the least amount of rainfall. This resulted because during that time the annual average maximum temperature was really high in the entire district so the precipitation was really poor.

5.4 Data Analysis of humidity

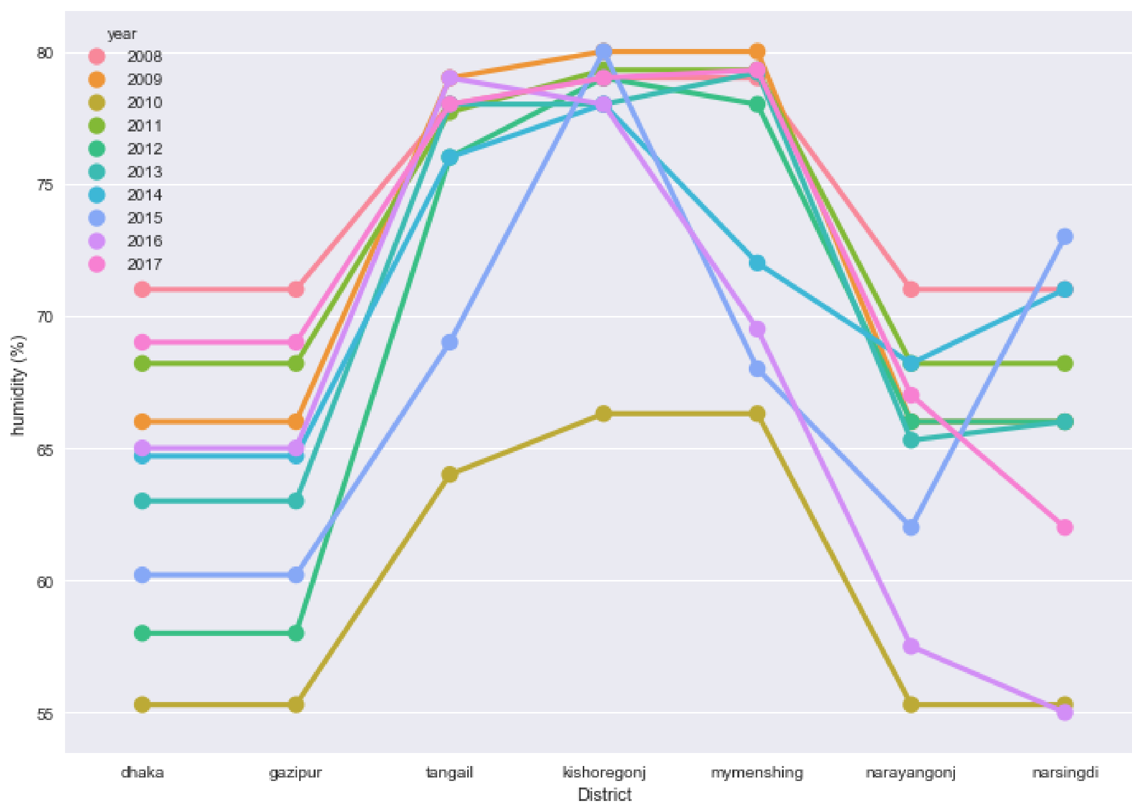

Figure 5.4: Analysis of Average Humidity

In figure 5.4, shows the result analysis of average humidity from the year 2008 to 2017 in Dhaka, Gazipur, Narayangonj, Tangail, Kishoregonj, Mymenshing, and Narsingdi district. In the horizontal axis, we have set districts (Agricultural Zone-28) and in the vertical axis, we have set average humidity (%).In 2010 the humidity was very low in the entire districts. The average humidity was about 55% to 67%, where Dhaka, GazipurNarayangonj, and Narsingdi were stuck into 55%. In 2009, the percentage of humidity got the peak amount. But it desperately goes down until 2010. Again, the humidity increases every year about 5%. Kishoregonj district is the place, where the humidity is always high and because of the excessive humidity, this place rains a lot as well.

5.5Result Analysis of Aus Rice

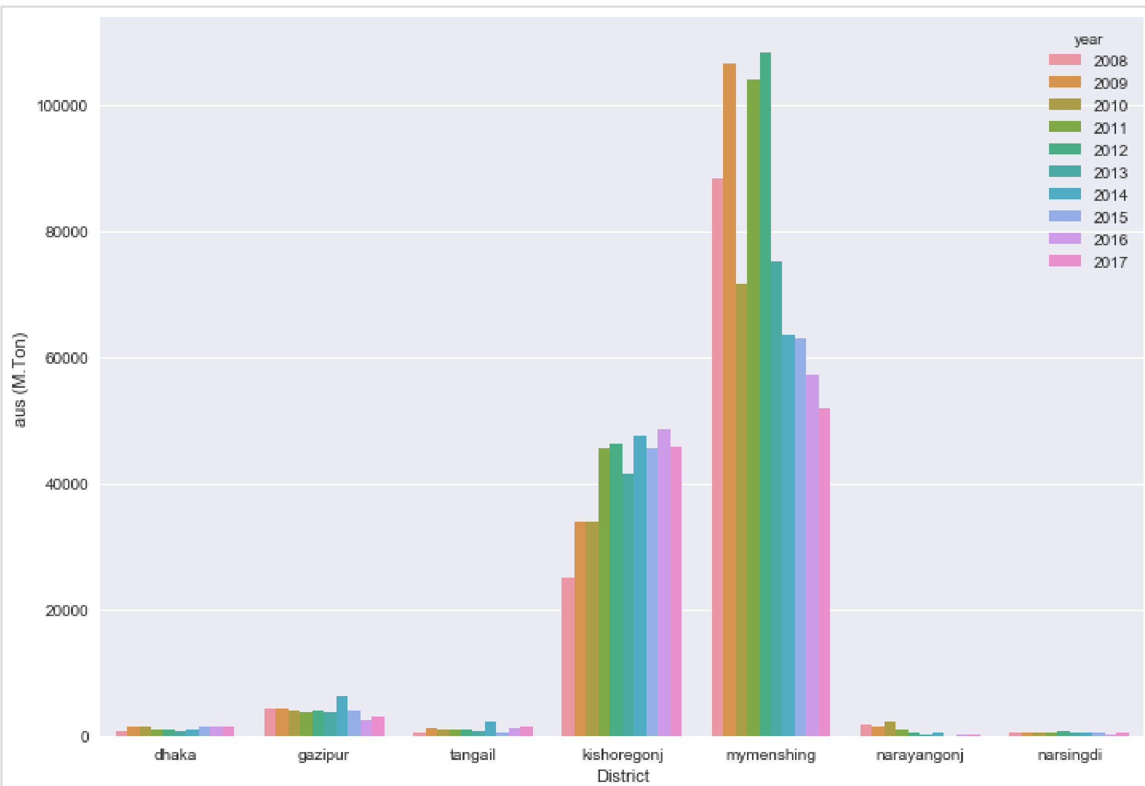

Figure 5.5.1: Analysis of crop production (Aus rice)

In figure 5.5.1, displays the result analysis of production of Aus rice from year 2008 to 2017 in Dhaka, Gazipur, Narayangonj, Tangail, Kishoregonj, Mymenshing and Narsingdi district .In the horizontal axis, we have set districts (Agricultural Zone-28) and in the vertical axis, we have set Aus rice production (M.ton).As we can see from the above figure Mymenshing is the leading district where Aus rice produced the most and on the other hand Tangail, Narsingdi, Narayangonj, Dhaka performed the poorest. In 2012 Mymenshing had the highest crop which was about 110000 M.ton. The reason behind this massive growth rate because of extreme chemical fertilizer which was found on this district over the year. Moreover, the humidity was higher (78%) but the precipitation was low comparing to other years. Furthermore the maximum temperature was quite decent but the minimum temperature for Mymenshing was upwards comparatively to other districts. Furthermore, in Mymenshing district there has huge number of non-calcareous dark grey floodplain soil, non-calcareous grey floodplain soil, deep red brown terrace soil, shallow red brown terrace soil and acid basin clay soil. As well as, we have found huge number of soil texture at Mymenshing area. That is the reason for peak amount of production level of Aus rice on that year. On the other hand if we look at the Narsingdi or Dhaka district we will notice that they both have low amount of Aus rice production. The reason behind is the chemical fertilizer was very low compare to Mymenshing district but one chemical fertilizer which is Triple Super Phosphate was higher for Dhaka district then Narsingdi district probably that's might be one reason that Dhaka district has quite better of crop production level then Narsingdi district. Furthermore if we look at the rainfall level of both the district Narsingdi had a bit on higher side where as Dhaka city rainfall gradually decreasing. Another thing we noticed that Dhaka city had quite higher level of temperature and humidity whereas Narsingdi had average temperature level.

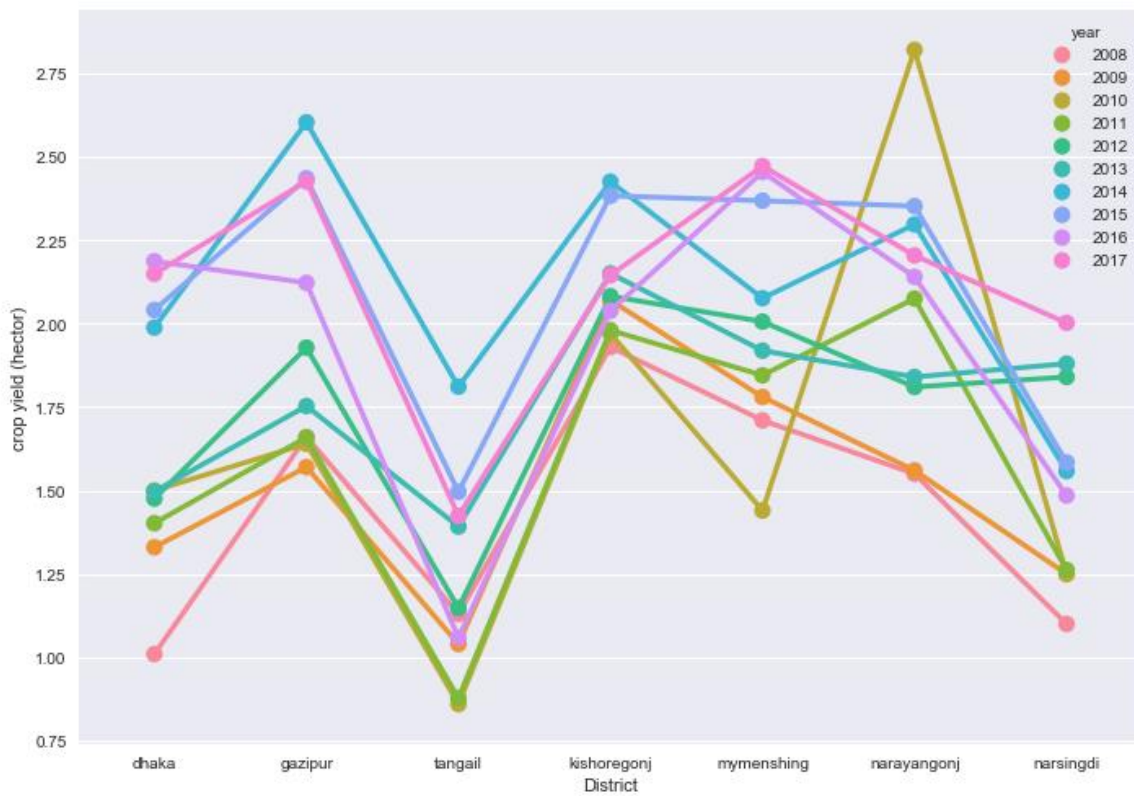

Figure 5.5.2: Analysis of crop yield (Aus rice)

In figure 5.5.2, describes the crop yield of Aus rice from the year 2008 to 2017 in Dhaka, Gazipur, Narayangonj, Tangail, Kishorgonj, Mymenshing, and Narsingdidistrict. This is the agricultural output of Ausrice.In the horizontal axis, we have set districts (Agricultural Zone-28) and in the vertical axis, we have set crop yield (hector) for Aus rice. There is a difference between production and crop yield. The reason because we are analyzing the crop yield is we can easily understand actually how much production occurs in a particular area. For measuring the crop yield we need plugin the value of production level and divide it with the area the production level. As from the above graph it is noticeable that although Mymenshing did the higher production level but comparatively to its area it didn't perform up to the mark. Although Gazipurhas very small amount of land compare to Mymenshing district but it did performed well from 2014 to 2017. Alike Mymenshing, Gazipur has huge number of inundation-land highland and crop yield rateis increasing linearly.

5.6 Result Analysis of Aman Rice

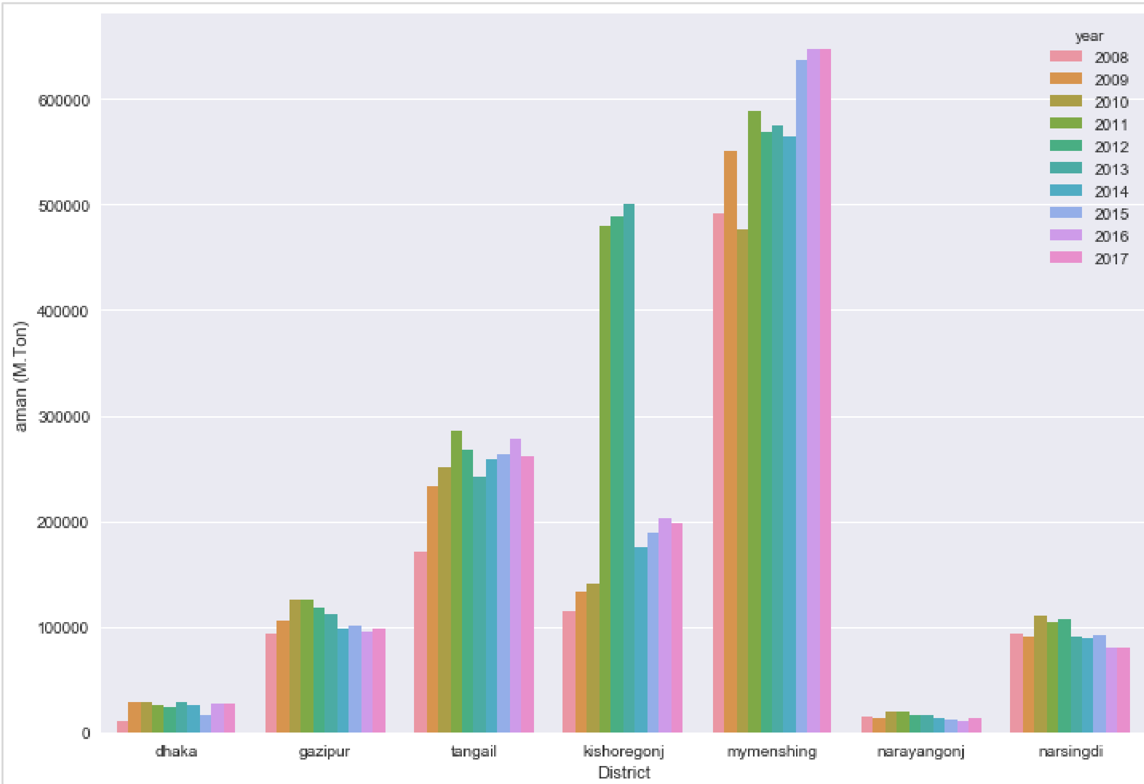

Figure 5.6.1: Analysis of crop production (Aman rice)

In figure 5.6.1, represents the result analysis of the production of Aman rice from the year 2008 to 2017 in Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymenshing, and Narsingdi district. In the horizontal axis, we have set districts (Agricultural Zone-28) and in the vertical axis, we have set Aman rice production (M.ton). As we can see from the above figure, Mymenshing is still the leading district where Aman rice produced the most. This is the same reason as Aus rice production. The pattern of Aus rice production and the production of Aman rice characteristic is pretty similar. From the above figure druing 2011 to 2013 Kishorgonj district seems to produce high amount of Amanrice. But if we look at the figure no 5.6.2 the crop yield increased from 2011 to 2017 actually. During 2014 to 2017 the land area of that particular district gradually decreases but on the other hand production actually increases which took the crop yield unit in maximum level. Moreover the main reason to improve its condition from past, was the changes of chemical fertilizer, soil, and land type. Kishorgonj has non-calcareous grey floodplain soil, non-calcareous dark grey floodplain soil, acid basin clay, non-calcareous alluvium, deep red- brown terrace soil, shallow red brown terrace soil which plays and important role to be capable of producing abundant crops. Moreover if we look at 2008 to 2010 we will notice all the 4 chemical components were very low but afterwards it tends to increases which actually help the Aman rice to expand its production level with proper soil consistency, soil moisture, soil reaction and soil texture. In 2008 because of excessive amount of rainfall and unstable atmosphere the production level decreases.

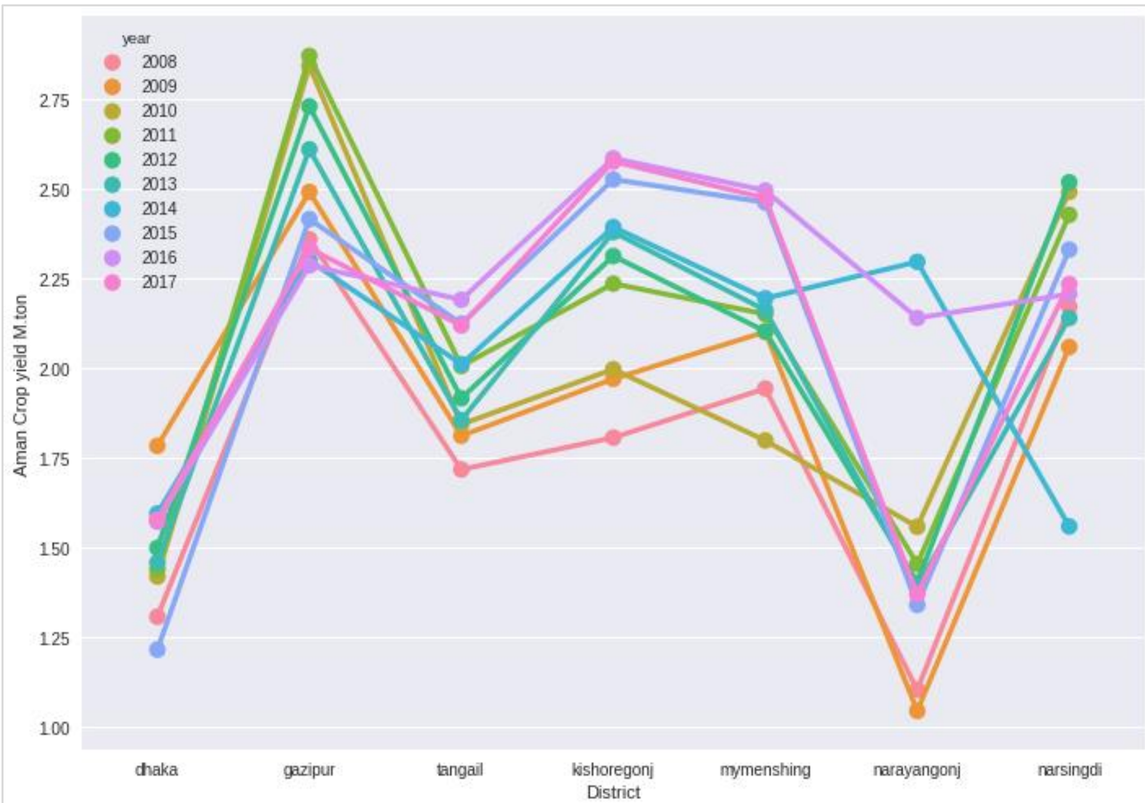

Figure 5.6.2: Analysis of crop yield (Aman rice)

In figure 5.6.2, describes the crop yield of Aman rice from the year 2008 to 2017 in Dhaka, Gazipur, Narayangonj, Tangail, Kishorgonj, Mymenshing, and Narsingdi district. This is the agricultural output of Aman rice .In the horizontal axis, we have set districts (Agricultural Zone-28) and in the vertical axis, we have set crop yield hector (metric ton) for Aman rice. From figure no 5.5.1 and figure no 5.6.1 we can easily see that Gazipur and Narsingdi performed really well. Land area plays an important role here for

Gazipurand Narsingdi. Both districts has same type of land which are inundation land medium highland,inundation land highland, inundationland low land,inundation land medium low land and miscellaneous land. One major drawback is due to lack of small area of land like other district those district cannot improve the yielding level. In conclusion it can be said that Gazipur and Narsingdi would be a really suitable place for Aus and Aman rice production.

5.7Result Analysis of Boro Rice

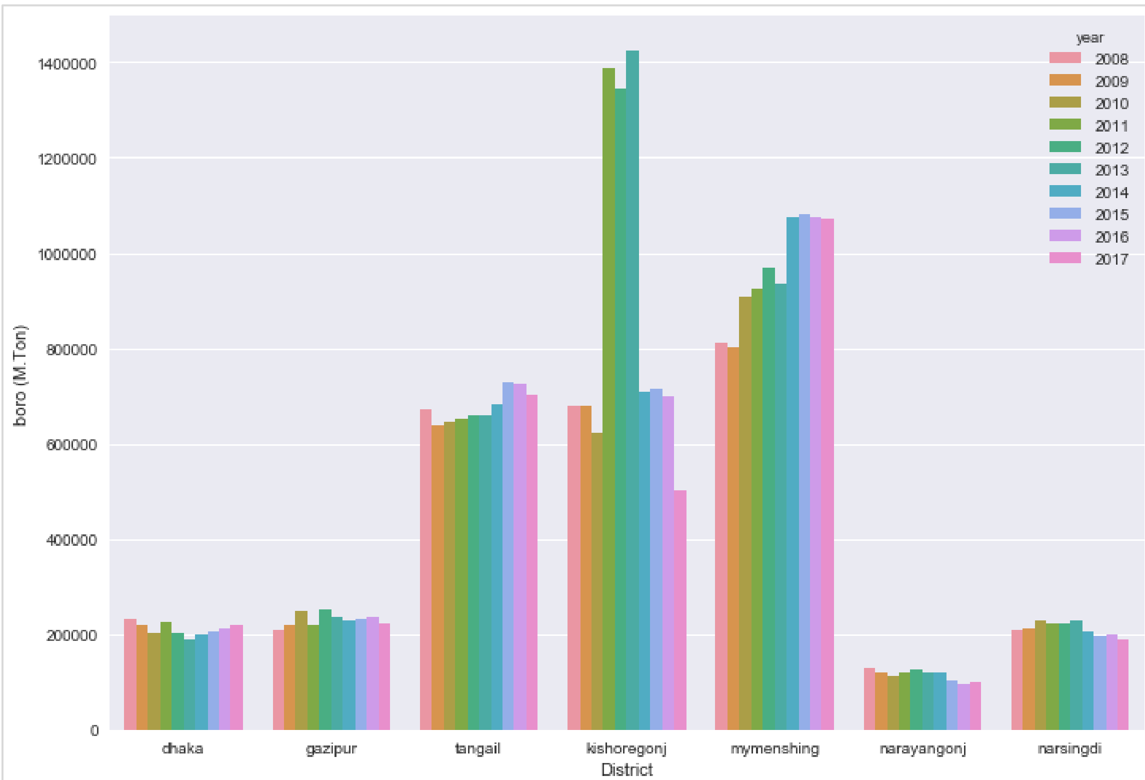

Figure 5.7.1: Analysis of crop production (Boro rice)

In figure 5.7.1, represents the result analysis of the production of Bororice from the year 2008 to 2017 in Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymenshing, and Narsingdi district. In the horizontal axis, we have set districts (Agricultural Zone-28) and in the vertical axis, we have set Boro rice production (metric ton). From previous production figure of Aus and Aman rice we noticed that Dhaka didn’t produce a good amount of yield rate. But in figure no 5.7.1 production rate of Boro in Dhaka increased significantly. Furthermore Dhaka has inundation land highland,inundation land medium low highland,inundation land low land and miscellaneous land which is consist of non-calcareous alluvium soil, acid basin soil, and calcareous dark grey floodplain soil with high concentration which better for Boro rice cultivation. Moreover, this tends to happened as a result of the weather condition which was not suitable for producing Aus and Aman rice, whereas that weather and chemical fertilizer suits Boro rice farming in Dhaka city.

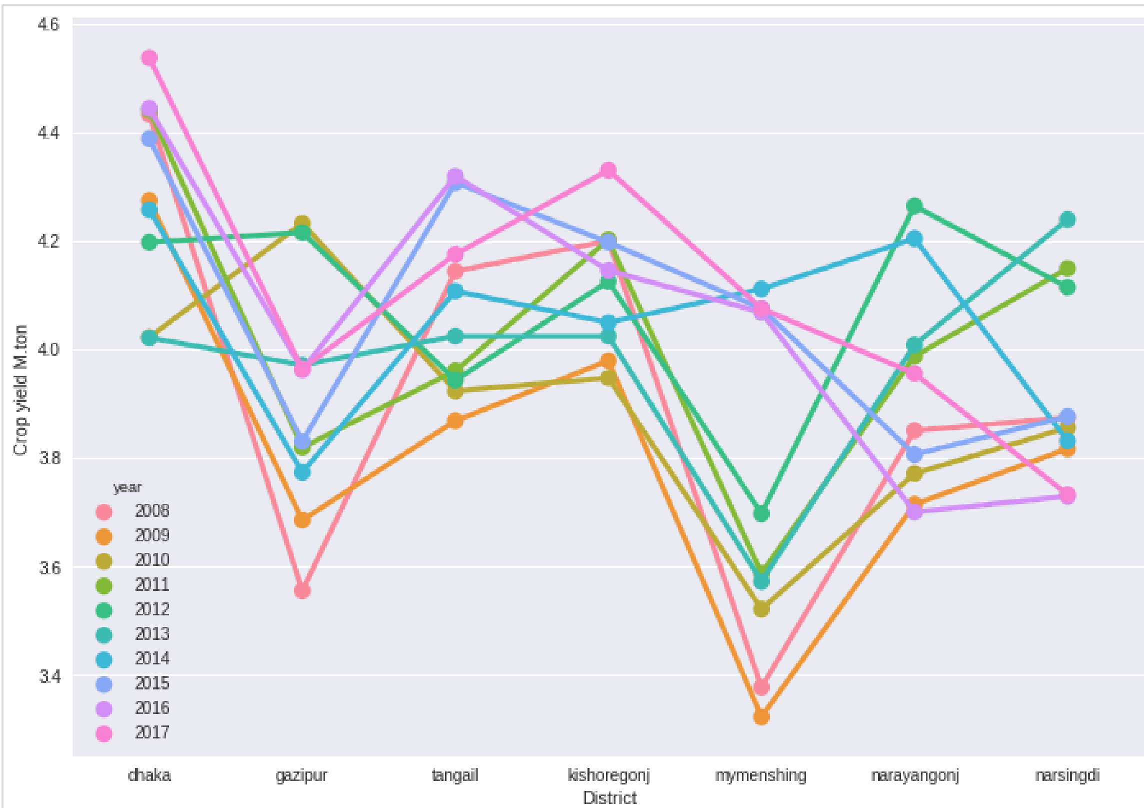

Figure 5.7.2: Analysis of crop yield (Boro rice)

In figure 5.7.2, describes the crop yield of Boro rice from the year 2008 to 2017 in Dhaka, Gazipur, Narayangonj, Tangail, Kishorgonj, Mymenshing, and Narsingdi district. This is the agricultural output of Bororice.In the horizontal axis, we have set districts (Agricultural Zone-28) and in the vertical axis, we have set crop yield hector (metric ton) for Boro rice. From the above figure we can see Dhaka has the highest amount of Boro rice yield and it was improving gradually. But in Mymenshing it is not suitable to cultivating Boro rice. Another thing we notice from the above figure Mymenshing didn’t performed well during 2008 to 2013 but eventually the crop yield rate increased very nicely which proves that there is higher probability that afterwards Mymenshing may become suitable to grow Boro rice.

5.8Result Analysis of Jute

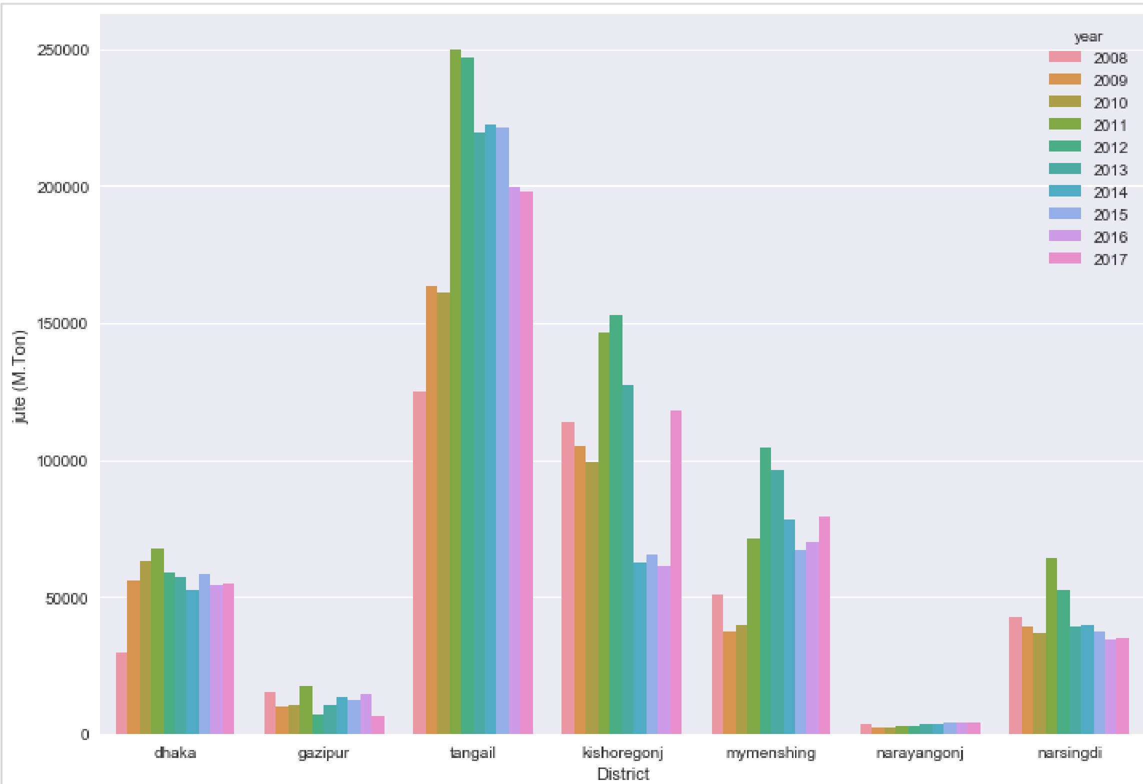

Figure 5.8.1: Analysis of crop production (Jute)

In figure 5.8.1, represents the result analysis of the production of Jutefrom the year 2008 to 2017 in Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymenshing, and Narsingdi district [10]. In the horizontal axis, we have set districts (Agricultural Zone-28) and in the vertical axis, we have set Jute rice production (metric ton). As we can see from the above figure that Tangail is the leading district with highest production whereas Mymenshing has low amount of yielding rate. Types of land found in Tangail are consists of different pattern of soil which are, non-calcareous alluvium, acid basin clay, calcareous grey floodplain soil, non-calcareous grey floodplain,non-calcareous grey floodplain, deep red-brown soil, deep grey terrace soil, grey valley soil. Reasoning for this highest rate of production in Tangail, is Jute is a thirsty plant and Tangail has the significant rainfall with average high temperature and humidity. In addition to these soil of Tangail has best soil moisture among other three types soil information. Moreover the fertilizers were high in Mymenshing and less soil moisture than soil texture which is good for rice production but not for fertility-exhausting plant like jute.

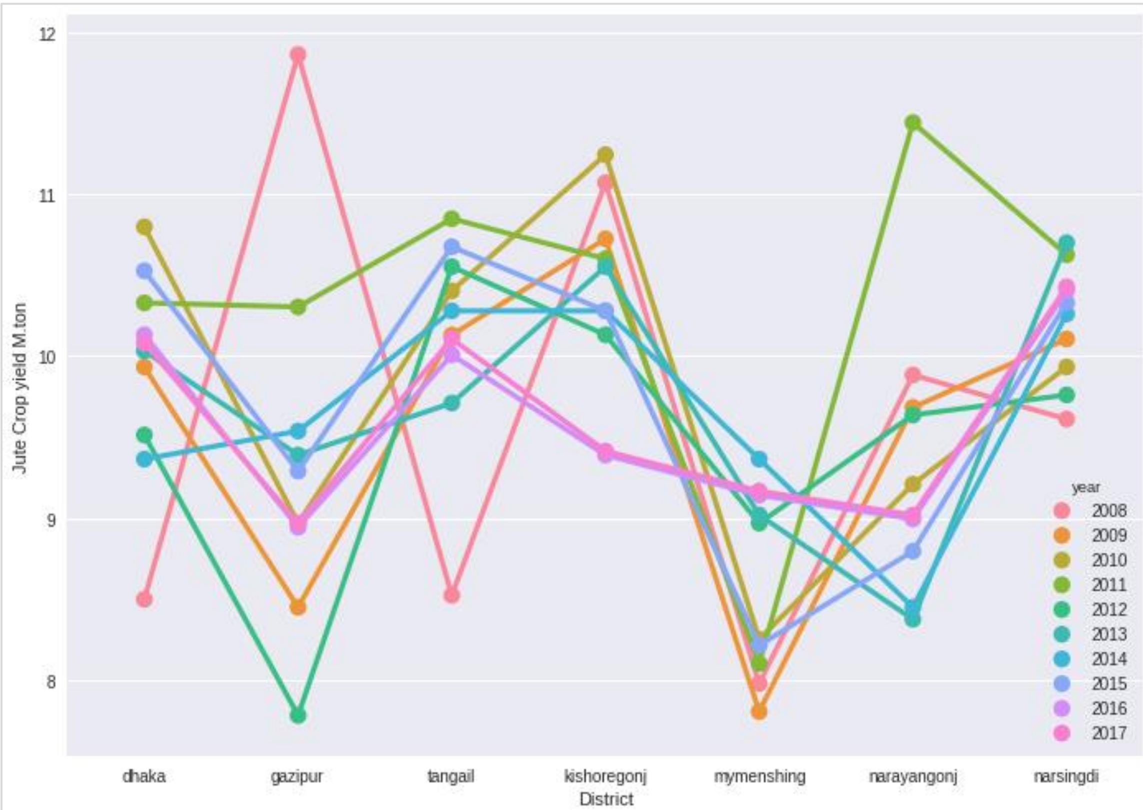

Figure 5.8.2: Analysis of crop yield (jute)

In figure 5.8.2, describes the crop yield of Jute from the year 2008 to 2017 in Dhaka, Gazipur, Narayangonj, Tangail, Kishorgonj, Mymenshing, and Narsingdi district. This is the agricultural output of Jute. In the horizontal axis, we have set districts (Agricultural Zone-28) and in the vertical axis, we have set crop yield hector (metric ton) for Jute. From the figure 5.8.1 Tangail has highest production but slightly decreased after 2015. For this in figure 5.8.2 Kishoregonj&Narsingdi has a good amount of yielding rate because of stable required environment, atmosphere, soil and fertilizer. Furthermore from the graph we can see that ratio of Gazipur rate over the year is more fluctuating. 2008 Gazipur has the highest amount of yielding rate (11.8M.ton) but it decreased drastically in 2009. In 2009 temperature was very low near to 11 degree Celsius which is not feasible for keeping the rate same as 2008.

5.9Result Analysis of wheat

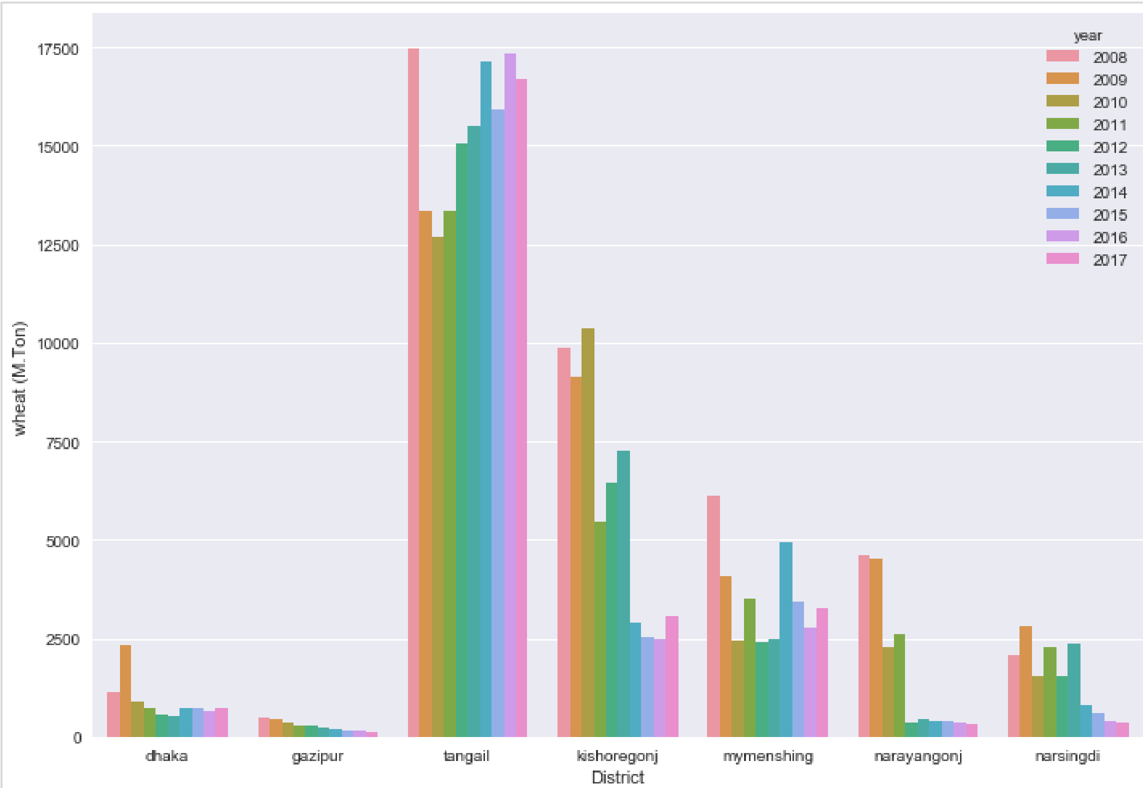

Figure 5.9.1: Analysis of crop production (wheat)

In figure 5.9.1, represents the result analysis of the production of wheat from the year 2008 to 2017 in Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymenshing, and Narsingdi district. In the horizontal axis, we have set districts (Agricultural Zone-28) and in the vertical axis, we have set wheat production (metric ton). From above figure Tangail produced highest production rather than other districts. Tangail has average fertilizers with a bit high amount of urea, minimum temperature below average, high humidity and an unstable rainfall which created a very less affect in production of wheat. For these events Tangail always have a moderated amount of production among all districts. On the other hand from figure 5.9.1, we can see Gazipur couldn’t produce an average production of wheat. Reasons of this can be soil type difference of Gazipur. Gazipur always has deep red brown terrace soil, shallow red brown terrace soil, acid basin clay, grey valley soil, shallow grey soil, and calcareous alluvium soil. Moreover the soil texture is higher than any other soil information whereas wheat production increases in more in those soil which has more soil moisture in common.

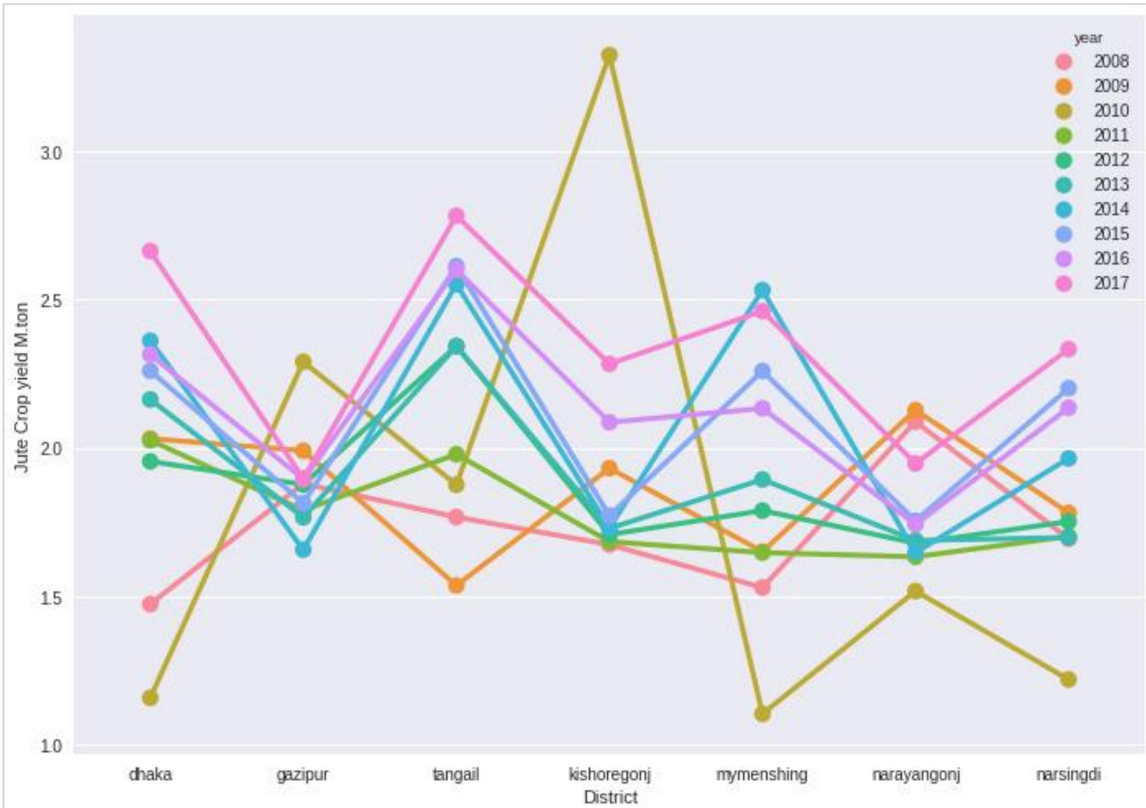

Figure 5.9.2: Analysis of crop yield (wheat)

In figure 5.9.2, describes the crop yield of Wheat from the year 2008 to 2017 in Dhaka, Gazipur, Narayangonj, Tangail, Kishoregonj, Mymenshing, and Narsingdi district. This is the agricultural output of Wheat. In the horizontal axis, we have set districts (Agricultural Zone-28) and in the vertical axis, we have set crop yield hector (metric ton) for Jute. From the above figure, we can see almost most of the district the jute yield was about 1.5 hector (metric ton) to 2.5 hector (metric ton). In 2010, Kishoregonj district had the highest amount of wheat yield which was more than 3.0 hector (metric ton) but eventually it falls down. During 2010, there was lowest humidity, average temperature and average precipitation, which can be effect on highest growth on wheat production of 2010 from all the districts.

5.10Result Analysis of Potato

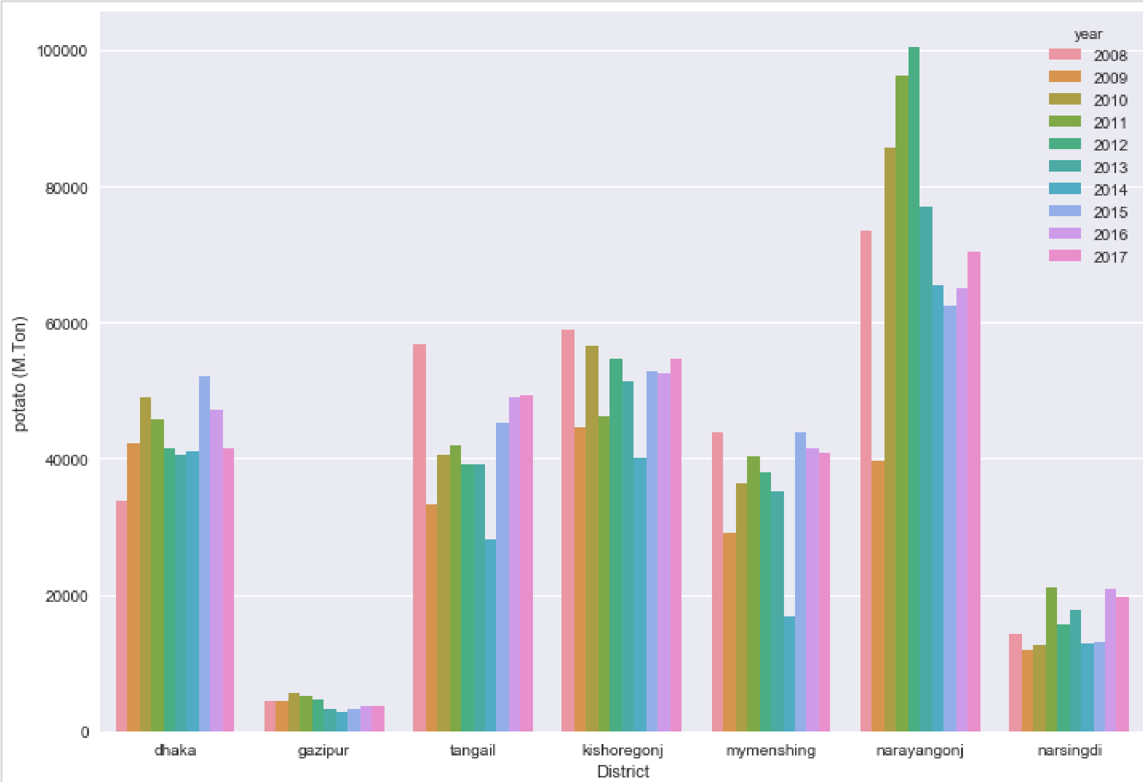

Figure 5.10.1: Analysis of crop production (potato)

In figure 5.10.1, represents the result analysis of the production of potato from the year 2008 to 2017 in Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymenshing, and Narsingdi district. In the horizontal axis, we have set districts (Agricultural Zone-28) and in the vertical axis, we have set potato production (metric ton). From figure 5.10.1 we can say that Narayanganj has the highest growth rate. The soil type of Narayanganj has inundation land medium low land, inundation land low land, and inundation land medium highland and miscellaneous land. These types land contains non-calcareous dark grey floodplain soil, non-calcareous grey floodplain soil, deep red brown terrace soil, shallow red brown terrace soil, acid basin clay and non-calcareous alluvium. These land proportion and soil proportion play a big role to produce potato as it’s a starchy and tuberous crop. Due to decreasing amount of maximum temperature, low amount of diammonium Phosphate and humidity, increasing rate of minimum temperature from 2008 potato growth gets higher till 2017. Gazipur, Tangail and Kishoregonjhave more or less same kind of production because of the similar environment, atmosphereand soil features.

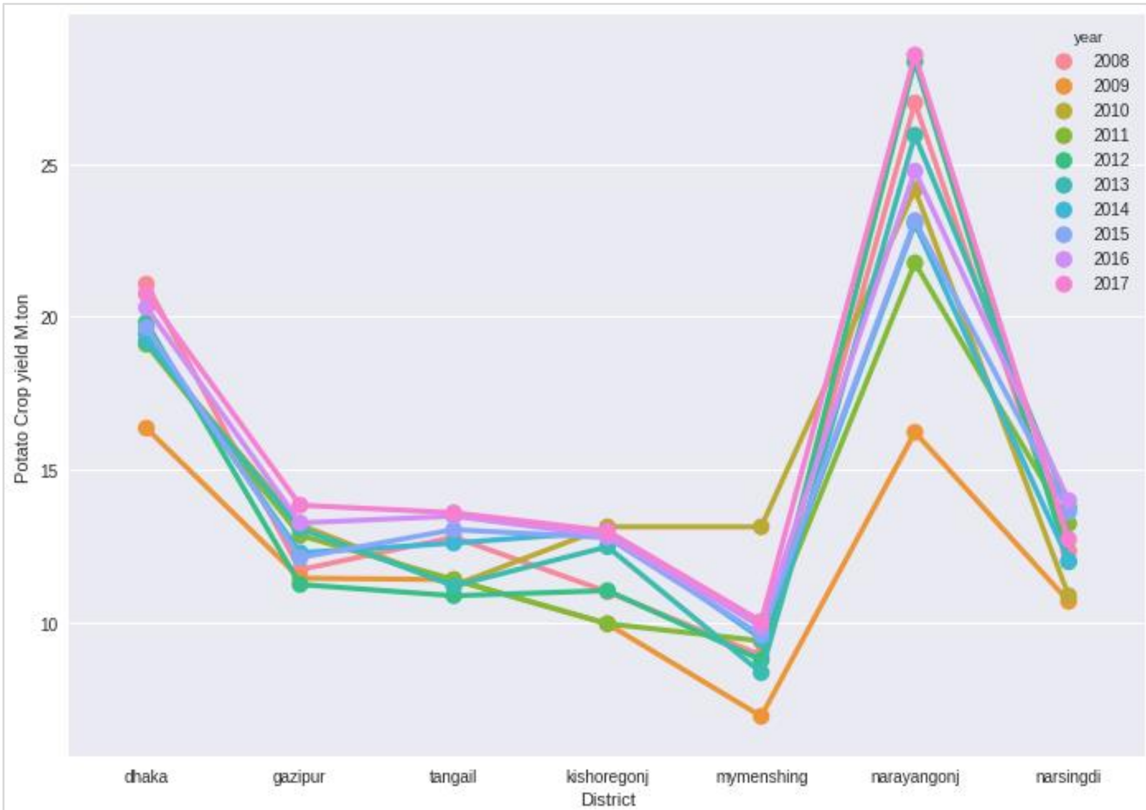

Figure 5.10.2: Analysis of crop yield (potato)

In figure 5.10.2, displays the crop yield of potato from the year 2008 to 2017 in Dhaka, Gazipur, Narayangonj, Tangail, Kishoregonj, Mymenshing, and Narsingdi district. This is the agricultural output of potato. In the horizontal axis, we have set districts (Agricultural Zone-28) and in the vertical axis, we have set crop yield hector (metric ton) for Potato. From the above figure we can see the graph of Narayangonj tends to go higher for potato yielding and as well as Dhaka city also in the 2^{nd} position for yielding potato. As Dhaka have average rainfall, average humidity, average temperature and average fertilizers. Moreover both Dhaka and Narayanganj has same proportion of soil consistency which leads their potato production in higher amount. Gazipur, tangail, Kishorgonj and Mymenshing tends to have the similar amount of potato growth over the past decades.

5.11Result Analysis of chemical fertilizer

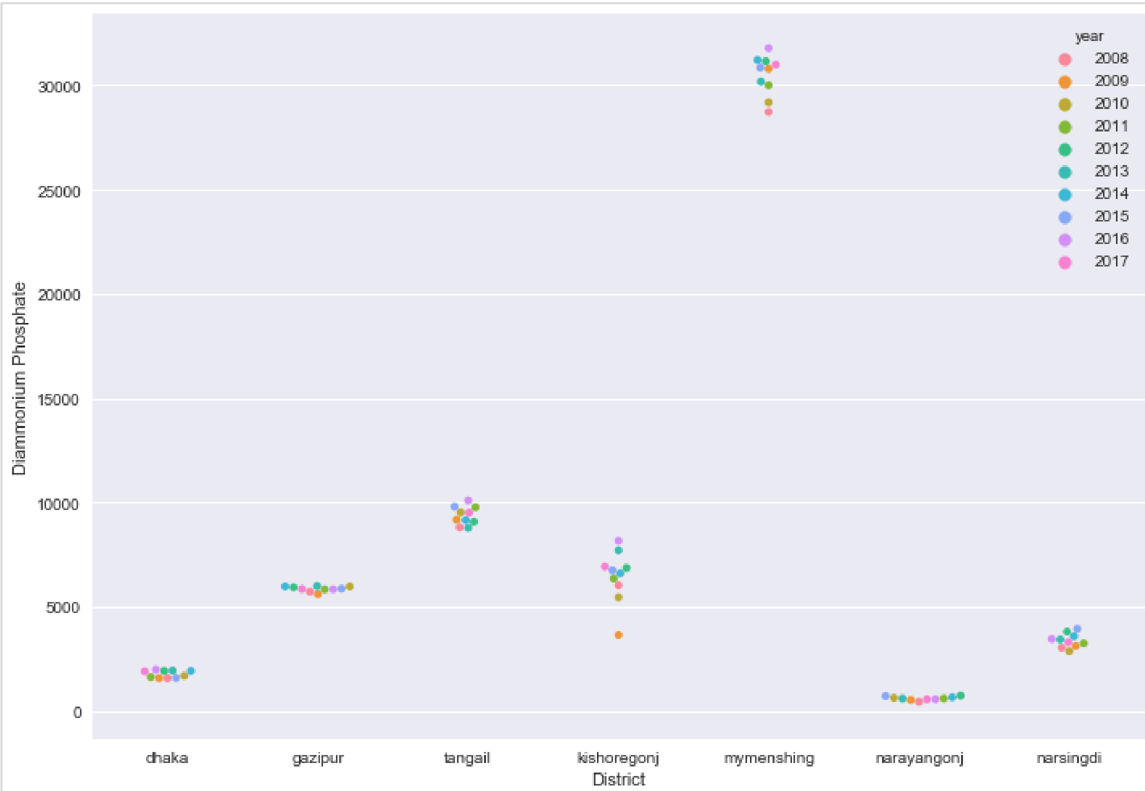

Figure 5.11.1: Analysis of fertilizer (Diammonium Phosphate)

In figure 5.11.1, displays the analysis of Diammonium Phosphate (chemical fertilizer) from the year 2008 to 2017 in Dhaka, Gazipur, Narayangonj, Tangail, Kishoregonj, Mymenshing, and Narsingdi district. Diammonium Phosphate is one of a series of water-solubleammoniumphosphatesalts that can be produced when ammonia reacts with phosphoric acid. When we applied Diammonium Phosphate as plant food, it increases the soilpH. In the horizontal axis, we have set districts (Agricultural Zone-28) and in the vertical axis, we have set value of DiammoniumPhosphate.This figure shows that,Mymenshing holds on the highest amount of Diammonium Phosphate fertilizer, which isaround 30000 units. Rest of the 6 districts Diammonium Phosphate fertilizer is moreover inbetween 3000 to 10000 unit.

Figure 5.11.2: Analysis of fertilizer (Urea)

In figure 5.11.2, displays the analysis of Urea (chemical fertilizer) from the year 2008 to 2017 in Dhaka, Gazipur, Narayangonj, Tangail, Kishoregonj, Mymenshing, and Narsingdi district. Urea, also known as carbamide, is an organic compound with chemical formula $CO(NH_2)_2$.This amide has two $–NH_2$ groups joined by a carbonyl (C=O) functional group.More than 90% of world industrial production of urea is destined for use as a nitrogen-release fertilizer. Urea has the highest nitrogen content of all solid nitrogenous fertilizers in common use.In the horizontal axis of this figure, we have set districts (Agricultural Zone-28) and in the vertical axis, we have set value of Urea.This figure shows that Mymenshing again holds on the highest amount of urea fertilizer which is about 120000 units. However, in Kishoregonj the quantity of urea fertilizer is falling drastically. Rest of the districts fertilizer is static.

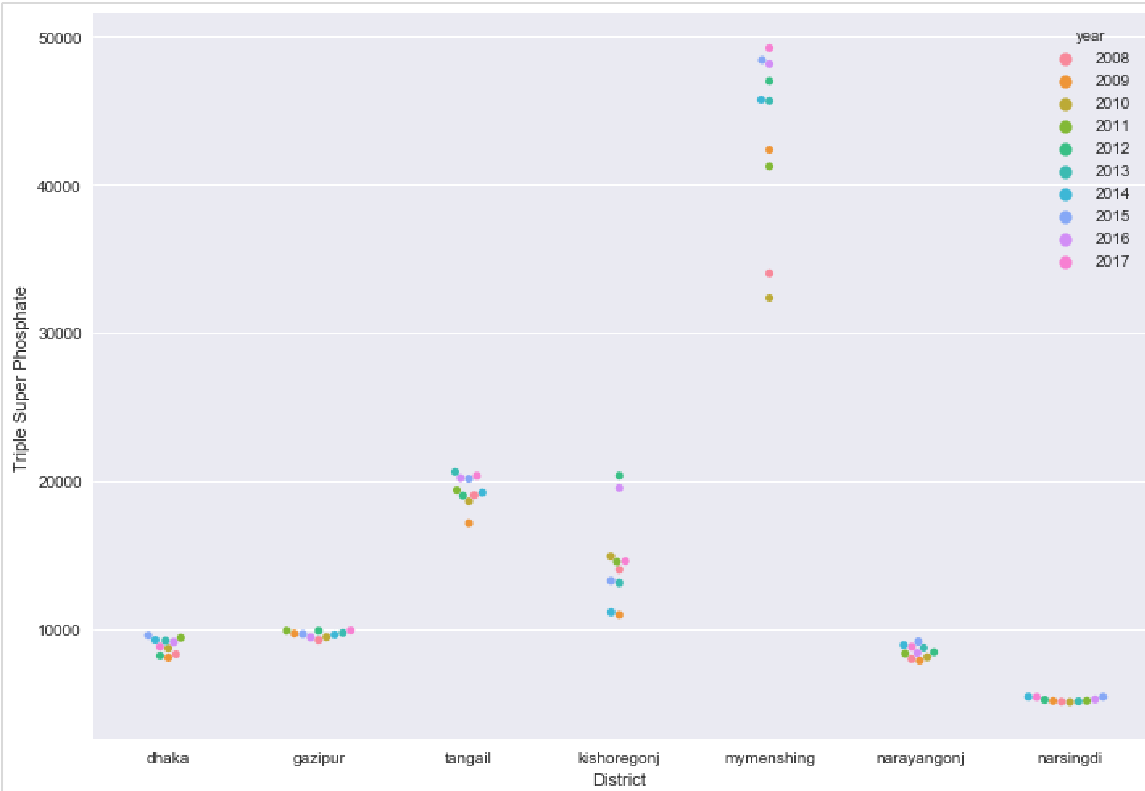

Figure 5.11.3: Analysis of fertilizer (Triple Super Phosphate)

In figure 5.11.3, displays the analysis of Triple Super Phosphate(chemical fertilizer) from the year 2008 to 2017 in Dhaka, Gazipur, Narayangonj, Tangail, Kishoregonj, Mymenshing, and Narsingdidistrict.Triple superphosphate (TSP) was one of the first high-analysis phosphorus (P) fertilizers that became widely used in the 20th century. Technically, it is known as calcium dihydrogen phosphate and as monocalcium phosphate, [Ca(H2PO4)2.H2O].In the horizontal axis of this figure, we have set districts (Agricultural Zone-28) and in the vertical axis, we have set value of MP.This figure shows that Mymenshing again holds on the highest amount of Triple Super Phosphate fertilizer. But during 2010 the amount of Triple Super Phosphate fertilizer was in fewer amounts but afterwards it starts to pull gradually. Rest of the districts fertilizer is hasn't changed much.

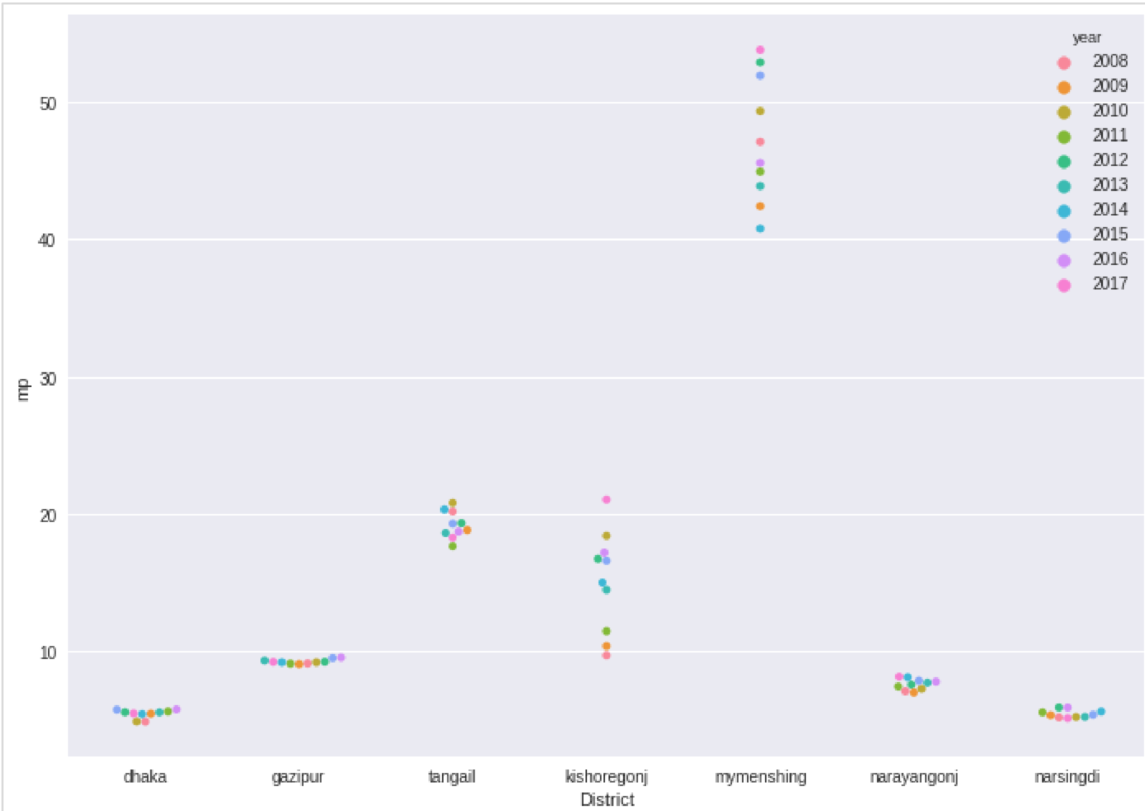

Figure 5.11.4: Analysis of fertilizer (mp)

In figure 5.11.4, displays the analysis of MP (chemical fertilizer) from the year 2008 to 2017 in Dhaka, Gazipur, Narayangonj, Tangail, Kishoregonj, Mymenshing, and Narsingdidistrict.In the horizontal axis of this figure, we have set districts (Agricultural Zone-28) and in the vertical axis, we have set value of Triple Super Phosphate. This is the result analysis of mp fertilizer from year 2008 to 2017 in Dhaka, Gazipur, Narayanganj, Tangail, Kishorgonj, Mymenshing and Narsingdidistrict.This figure again shows that Mymenshing again holds on the highest amount of mp fertilizer. So in conclusion, it is proved that all of the 4 essential fertilizers which is important for yield production are maximum situated in Mymenshing district.

5.12.1 Result Analysis of types of land

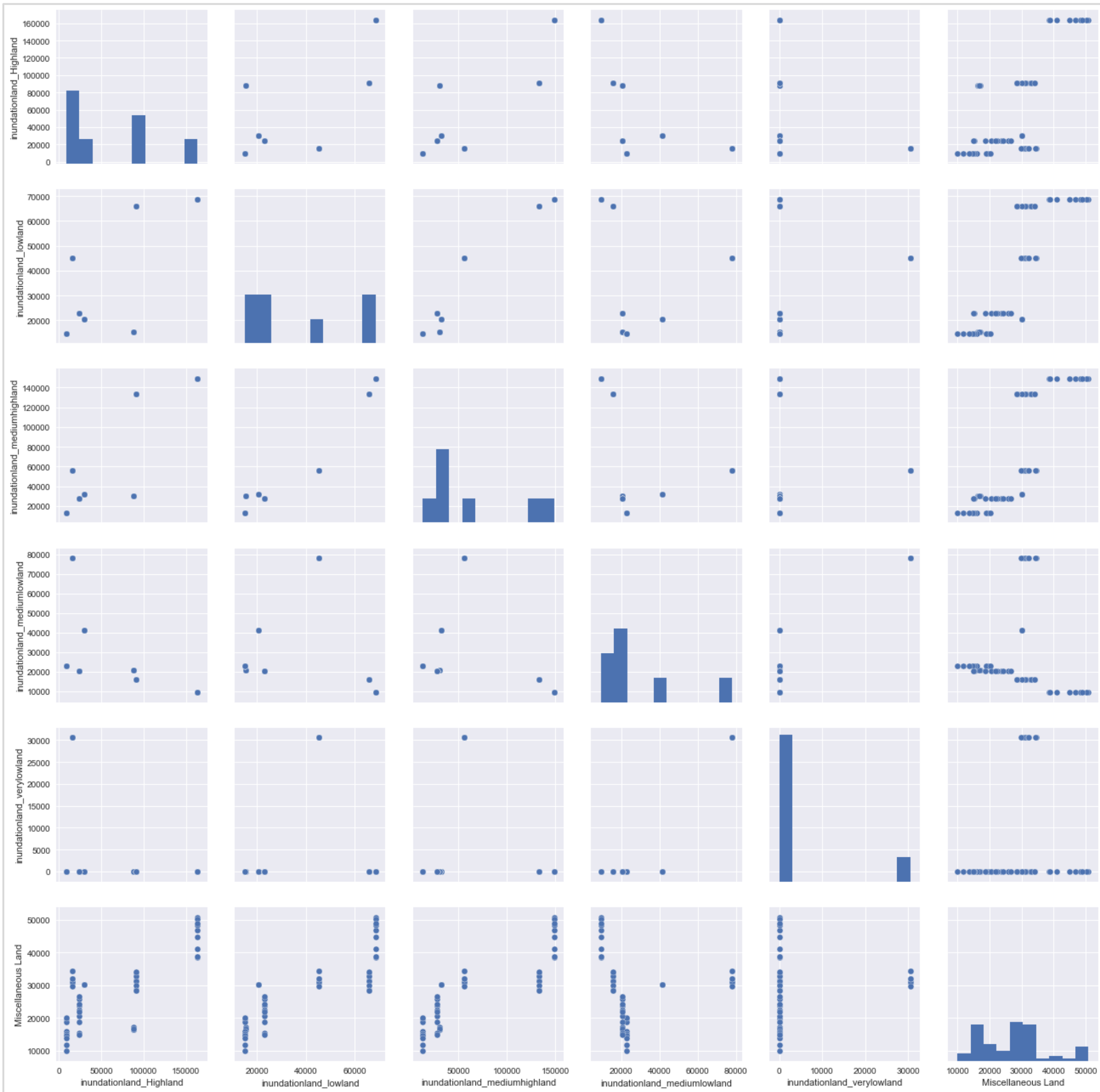

Figure 5.12.1: Analysis of pair plotting (land)

In figure 5.12.1, shows the pair plot among types of land (Inundation-land highland, Inundation-land medium highland, Inundation-land medium low land, Inundation-land low land, Inundation-land very low land, miscellaneous land). Basically, pairs plot forms on two basic figures, the histogram and the scatter plot. The histogram on the diagonal allows us to see the distribution of a single variable while the scatter plots on the upper and lower triangles show the relationship between two variables. In figure wecan see that

Inundation-land low land and miscellaneous land positively correlated. It is showing that in the place where the amount of Inundation-land low land is higher there having a higher amount of miscellaneous land. On the other hand, we can see that Inundation-land very low land and miscellaneous land negatively correlated. So, it is representing that in the place where the amount of Inundation-land very low land is higher there having a lower amount of miscellaneous land. Furthermore, from the histograms, we can see that the Inundation-land highland and Inundation-land very low land are heavily right-skewed.

5.12.2 Result Analysis of pair plot of soil type

Figure 5.12.2: Analysis of pair plotting (soil)

In figure 5.12.2, shows the pair plot among types of soil (Calcareous Alluvium, Non-calcareous Alluvium, Acid Basin Clay, Calcareous brown floodplain soil, Calcareous grey floodplain soil, Calcareous dark grey floodplain soil). Basically, pairs plot forms on two basic figures, the histogram and the scatter plot. The histogram on the diagonal allows us to see the distribution of a single variable while the scatter plots on the upper and lower triangles show the relationship between two variables. In figure, we can see that in 2008, there has the highest amount of Non-Calcareous Alluvium soil and in 2014; there has the highest amount of Acid Basin Clay. Moreover, we can see that Calcareous Alluvium and Calcareous grey floodplain soil have the absence of linear relation. But we cannot infer about the nonlinearities. In addition, Calcareous Alluvium and Calcareous dark grey floodplain soil have the absence of linear relation. Furthermore, from the histograms, we can see that the Calcareous Alluvium Calcareous brown floodplain soil, Calcareous grey floodplain soil, and Calcareous dark grey floodplain soil are heavily right-skewed.

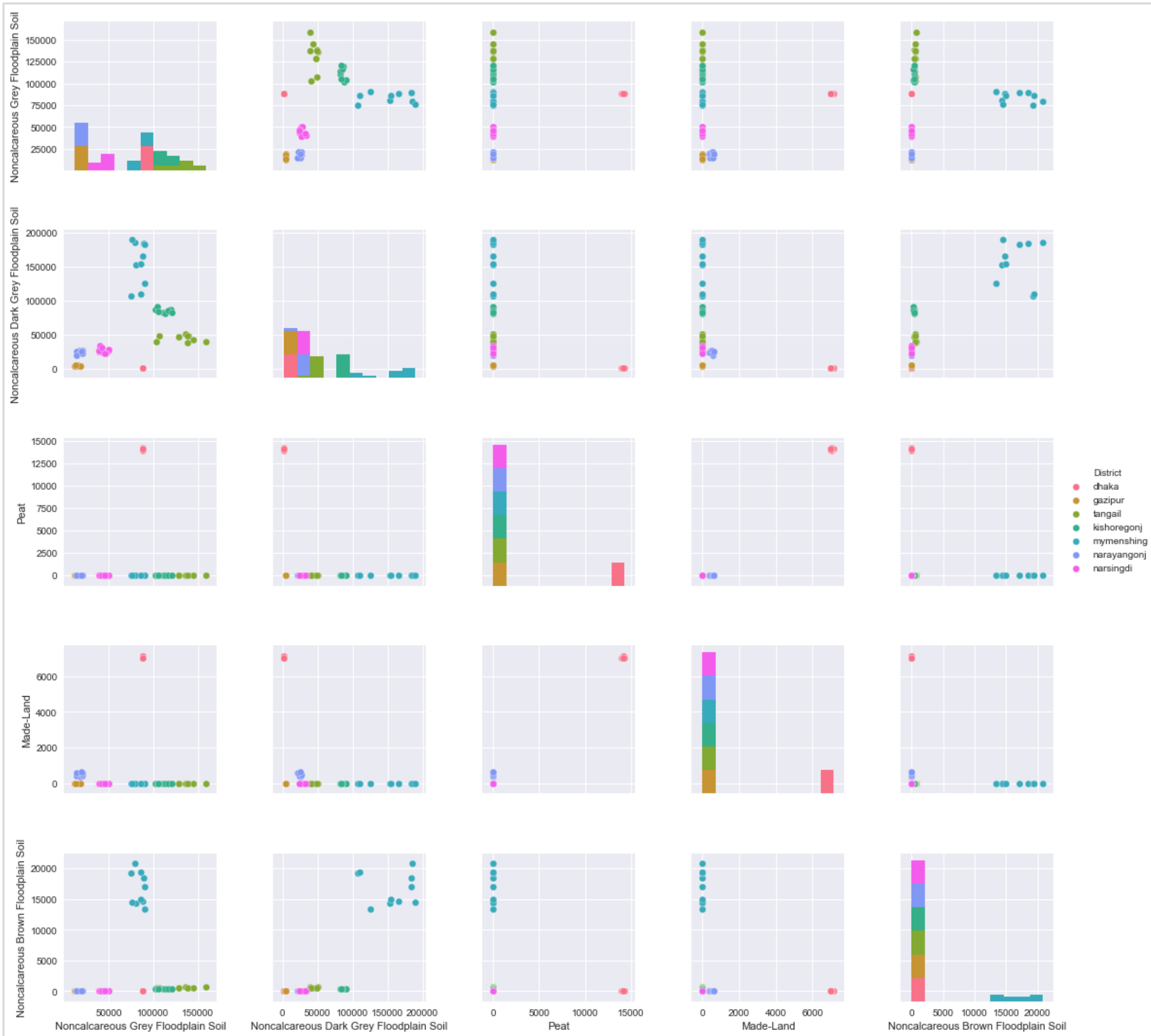

Figure 5.12.3: Analysis of pair plotting (soil)

In figure 5.12.3, shows the pair plot among types of soil (Non-calcareous grey floodplain soil, Non-calcareous dark grey floodplain soil, Peat, Made-land, Non-calcareous brown floodplain soil).In figure, we can see that Non-calcareous grey floodplain soil and Made-land negatively correlated. Also, Non-calcareous grey floodplain soil and Peat negatively correlated. Moreover, we can see that Peat and Made-land have the absence of linear relation. But we cannot infer about the nonlinearities. Furthermore, from the histograms, we can see that Non-calcareous grey floodplain soil, Non-calcareous dark grey floodplain soil, Peat, Made-land, Non-calcareous brown floodplain soils are heavily right-skewed. In addition, Mymenshing has highest amount of Non-calcareous dark grey floodplain soil and Dhaka has the highest amount of Peat.

Figure 5.12.4: Analysis of pair plotting (soil)

In figure 5.12.4, shows the pair plot among types of soil (Shallow red-brown terrace soil, deep red-brown terrace soil, Brown mottled terrace soil, Shallow grey terrace soil, deep grey terrace soil, grey valley soil, Brown hill soil, grey piedmont soil). From the above figure we can see that Shallow red-brown terrace soil and deep red-brown terrace soil positively correlated but it is showing that in Gazipur where the amount of Shallow red-brown terrace soil is higher there having a higher amount of deep red-brown terrace soil. On the other hand, Shallow red-brown terrace soil and Brown hill soil negatively related.

it is showing thatforGazipur, Tangail, Narayanganj, and Narsingdi, where the amount of Shallow red-brown terrace soil is higher there having a lower amount of deep red-brown terrace soil Moreover, we can see that Brown hill soil and grey piedmont soil have the absence of linear relation. However, we cannot infer about the nonlinearities.Furthermore, From the histograms, we can see that the Shallow red-brown terrace soil, deep red-brown terrace soil, Brown mottled terrace soil, Shallow grey terrace soil, deep grey terrace soil, grey valley soil, Brown hill soil, grey piedmont soil are heavily right-skewed.

5.13 Result Analysis of soil moisture and soil consistency

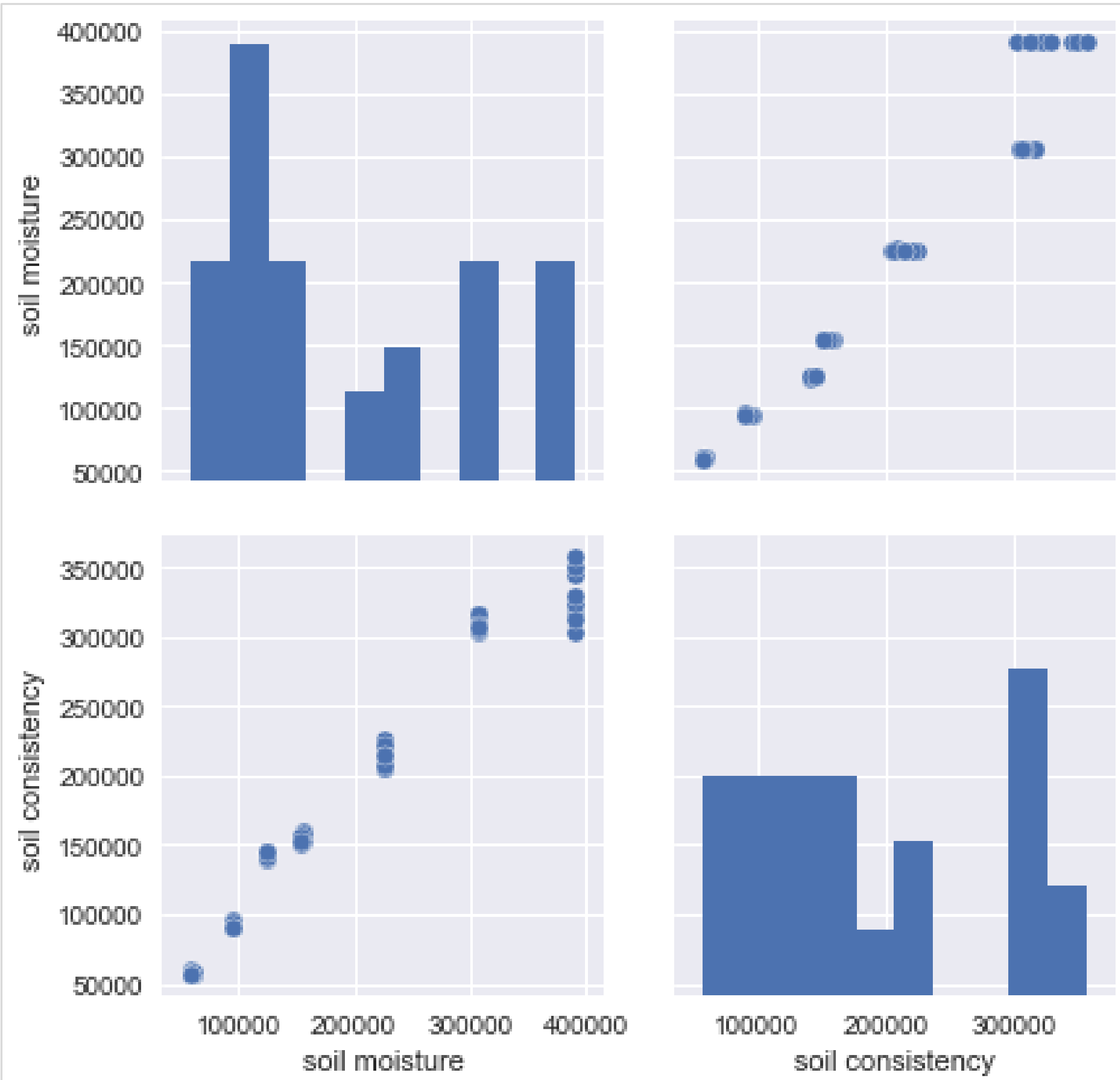

Figure 5.13: Analysis of pair plotting (Soil Consistency and Soil moisture)

In figure 5.13.1, shows the pair plotting between soil consistency and soil moisture. Pairs plot forms on two basic figures, the histogram and the scatter plot. The histogram on the diagonal allows us to see the distribution of a single variable while the scatter plots on the upper and lower triangles show the relationship between two variables. In this figure, we can see that soil consistency and soil moisture positively correlated. It is showing that in the place where the amount of soil moisture is higher there having a higher amount of soil consistency. Furthermore, from the histograms, we can see that soil moisture is right-skewed.

5.14Result Analysis of variable importance

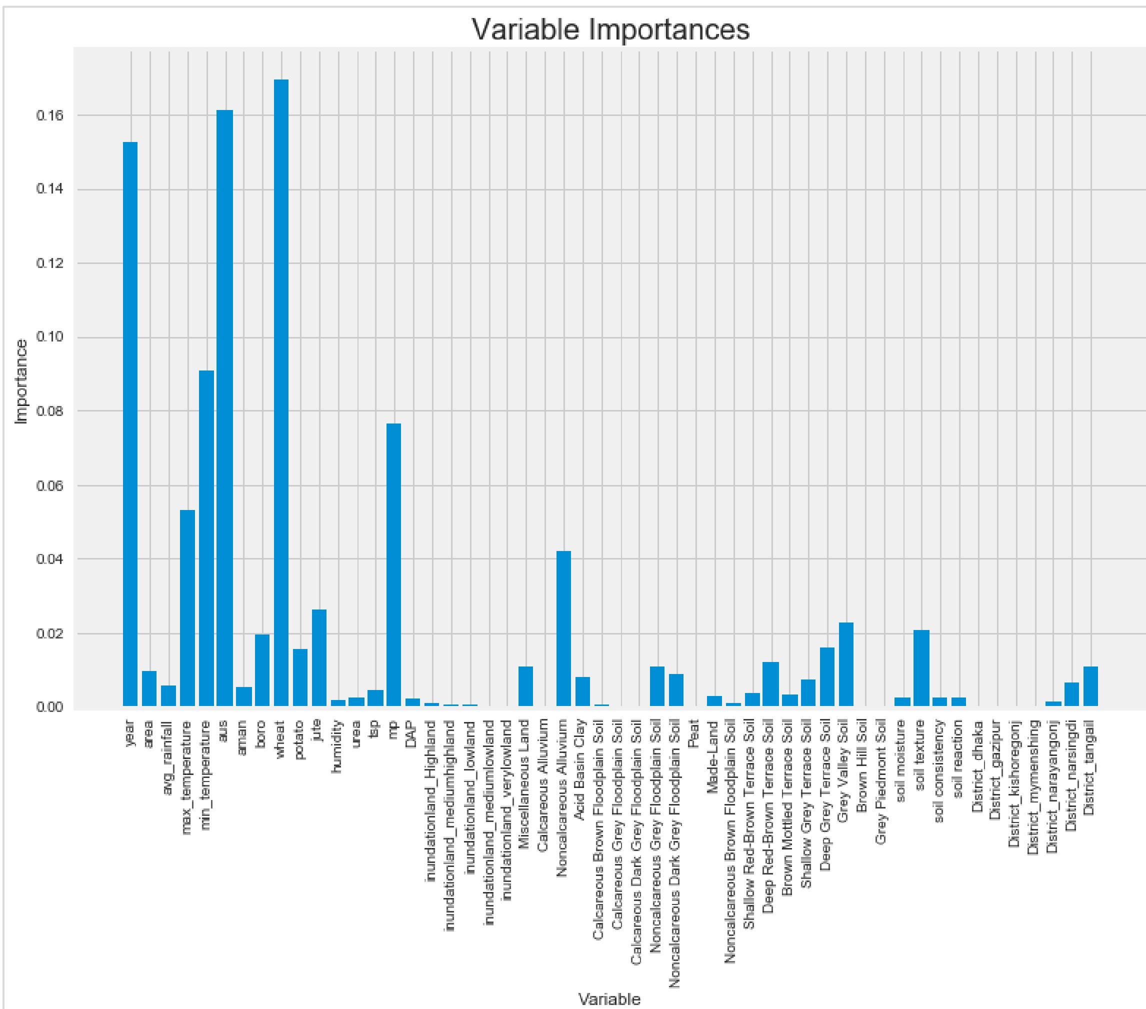

Figure 5.14: Analysis of variable importance

In figure 5.14, actually describes that during the time of our future prediction, which features, created the most importance. It is the statistical significance of each variable in the data with respect to its effect on the generated model. Therefore, from the diagram it is easily visible that meteorological data, chemical fertilizers and various soils are playing the vital role during our prediction.

5.15Result Analysis ofHeat-map of features

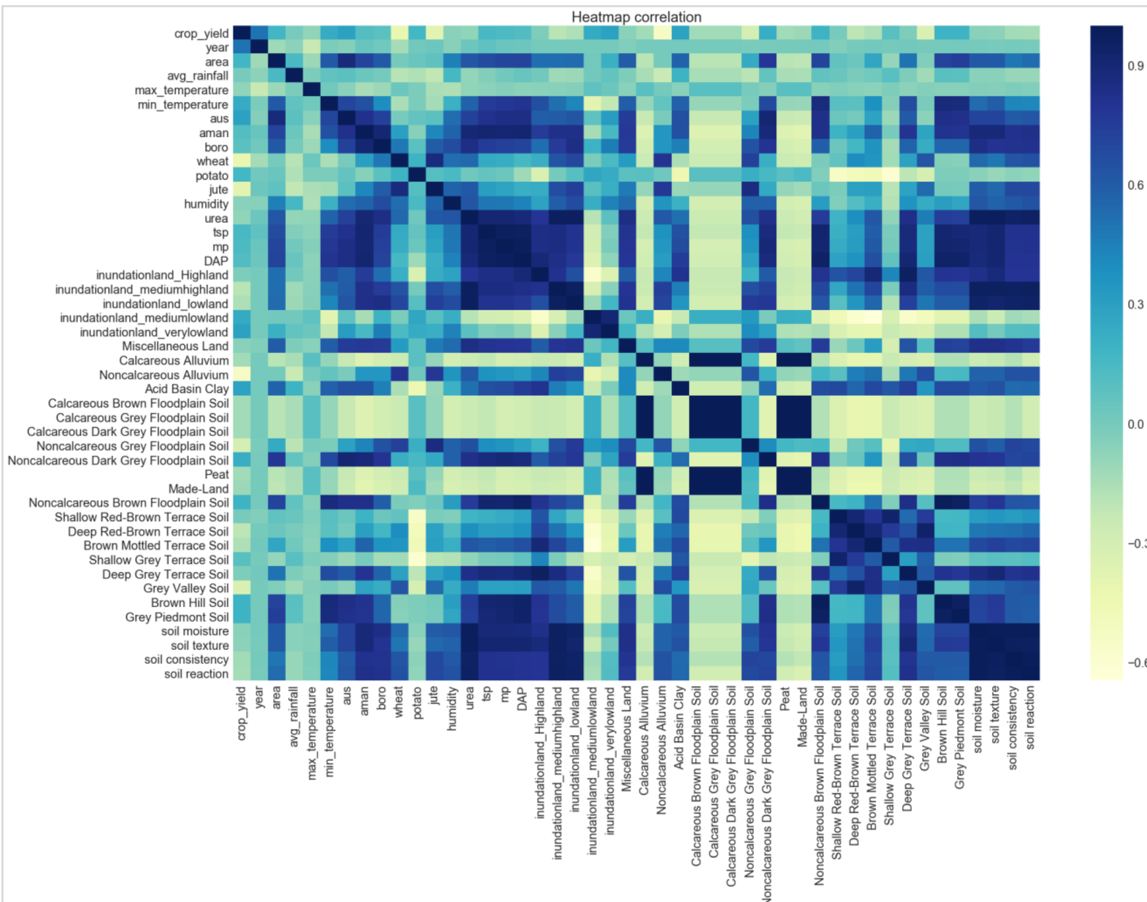

Figure 5.15: Heat-map of features

In figure 5.15, we can see a graphical representation of data where the individual 46 parameters contained in a matrix are represented as colors. Here, dark color represents strong co-relation between two variables.

5.16 Mean Squared Error (MSE)

MSE is the average of the squared error that is used as the loss function for least squares regression. So what it does it takes all the data points then it sum up those, of the square of the difference between the predicted and actual target variables, divided by the number of data point.

5.17 Root Mean Squared Error (RMSE)

RMSE is the square root of MSE. This means it is the standard deviation of the predicted errors. In other words how spread out the predicted errors are we can easily visualize through this loss function. This is frequently been used for measuring the differences between values predicted by a model.

5.18 Deep Neural Network system with Mean Squared Error

In our model we have used 2 loss functions for predicting the test score. Here we are showing how much accuracy we have achieved. We considered the 'tanh' activation function with 20 percent of validation set from the training set. Moreover we used 'sgd' optimizer and did 100 epochs to achieve a very good model. The reason behind doing this for making sure that our data is not get over fitted.

```
Loss:      0.05980263108556920 6
Val_loss:   0.0630701407790184
```

Figure 5.18.1:Loss and Validation loss result

```
14/14 [==============================] - 0s 147us/step
Test Score: 0.08804558217525482
56/56 [==============================] - 0s 172us/step
Train Score: 0.0601018661899226
```

Figure 5.18.2: Evaluating model in the train and test score result

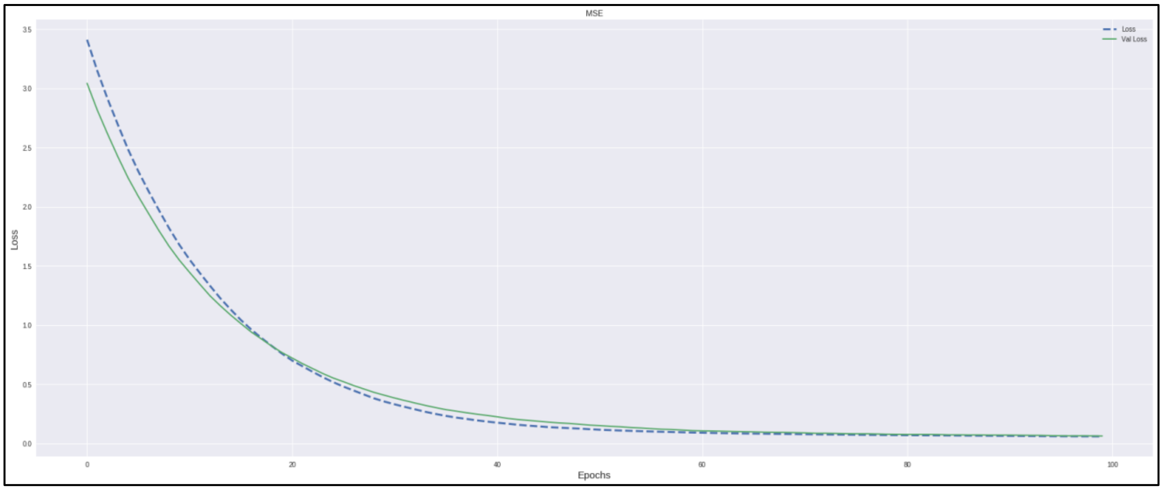

Figure5.18.3: Loss and Validation loss graph

Above graph shows that how precisely the loss is decreasing. This shows that our model learned perfectly. At the very beginning the loss was very high but gradually after every iteration the model learned itself and eventually the loss decreases significantly. After 100 epochs the loss was about .059 and the validation loss was about .063 which is impressive.

5.19 Result Analysis of prediction on Mymenshing district 2010

```
linha1 = np.array([2010,
                   49.829,
                   2.095,
                   30.7,
                   21.3,
                   71.548,
                   475.919,
                   908.157,
                   2.425,
                   36.304,
                   39.786,
                   66.3,
                   119.974,
                   32.347,
                   49.366,
                   29.182,
                   163.675,
                   148.743,
                   68.689,
                   9.442,
                   0,
                   50.861,
                   0,
                   9.667,
                   33.632,
                   0,
                   0,
                   0,
                   90.193,
                   125.976,
                   0,
                   13.458,
                   18.326,
                   30.396,
                   2.69,
                   1.285,
                   10.866,
                   7.307,
                   0.377,
                   17.103,
                   391.055,
                   382.617,
                   321.083,
                   333.083]).reshape(1,-1)
linha1 = sc_X.transform(linha1)
y_pred_mse_1 = classifier.predict(linha1)
print('Predict Test for Mymenshing (2010) : ')
print('Predicted value: ',y_pred_mse_1)
print('Real value: ','1.44')
```

```
Predict Test for Mymenshing (2010) :
Predicted value:  [[1.94144]]
Real value:  1.44
```

Figure 5.19: Prediction test for Mymenshing district 2010

This figure shows us how much accuracy we got at the end. Based on our training the model we try to evaluate how accurate our model gives us the prediction. So we considered genuine 46 Mymenshing district 2010 parameters value as the input and predicted our output. The actual value was about 1.44 and our model predicted 1.94 which is impressive. This was our first test so we decided to test same thing for other district as well.

5.20 Result Analysis of prediction on Narayangonj district 2015

```
linha1 = np.array([2015,
                   0.021,
                   1.633,
                   28.3,
                   15.2,
                   0.049,
                   12.516,
                   101.838,
                   0.397,
                   62.389,
                   3.777,
                   62,
                   17.964,
                   9.125,
                   7.798,
                   0.711,
                   9.076,
                   13.243,
                   14.834,
                   22.629,
                   0,
                   14.684,
                   0,
                   2.245,
                   2.915,
                   0,
                   0,
                   0,
                   0,
                   21.332,
                   27.424,
                   0,
                   0.504,
                   0,
                   2.314,
                   5.156,
                   0.021,
                   0.192,
                   0,
                   0.683,
                   60.077,
                   58.939,
                   57.837,
                   55.391]).reshape(1,-1)
linha1 = sc_X.transform(linha1)
y_pred_mse_2 = classifier.predict(linha1)
print('Predict Test for Narayangonj (2015) : ')
print('Predicted value: ',y_pred_mse_2)
print('Real value: ','2.352')
```

```
Predict Test for Narayangonj (2015) :
Predicted value:  [[2.1507156]]
Real value:  2.352
```

Figure 5.20: Prediction test for Narayangonj district 2015

Next we tried to predict the how much crop yield we may get in Narayangonj if we give the same 46 parameters input based on Narayangonj district 2015. The actual value was 2.352 and our model predicted 2.15 which is very close to our real value.

5.21 Result Analysis of prediction on Dhaka district 2014

```
linha1 = np.array([2014,
                   0.477,
                   1.399,
                   31.1,
                   14.2,
                   0.948,
                   25.676,
                   200.56,
                   0.71,
                   41.062,
                   52.2,
                   64.7,
                   35.321,9.233,
                   5.369,1.926,
                   30.118,32.245,20.539,
                   41.335,0,30.137,4.952,6.832,6.297,
                   7.703,26.322,28.199,87.938,1.896,14.095,
                   7.042,0,0,0,0,1,0,0,0,124.158,124.955,140.524,
                   112.11]).reshape(1,-1)
linha1 = sc_X.transform(linha1)
y_pred_mse_3 = classifier.predict(linha1)
print('Predict Test for Dhaka (2014) : ')
print('Predicted value: ',y_pred_mse_3)
print('Real value: ','1.988')
```

```
Predict Test for Dhaka (2014) :
Predicted value:  [[2.010409]]
Real value:  1.988
```

Figure 5.21: Prediction test for Dhaka district 2014

The above figure shows the result analysis of Dhaka district 2014. Our model predicted almost closer to the accurate result. The actual result was 1.988 and the predicted value was about 2.01. This one gives very much accurate result.

5.22 Result Analysis of prediction on Dhaka district 2008

```
linha1 = np.array([2008,
                   0.793,
                   2.385,
                   34.2,
                   12.5,
                   .804,
                   9.691,
                   233.939,
                   1.129,
                   33.679,
                   29.825,
                   8.262,
                   4.808,
                   1.573,
                   30.118,
                   32.245,
                   20.539,
                   41.335,
                   0,
                   30.137,
                   4.869,
                   6.787,
                   6.253,
                   7.691,
                   26.109,
                   28.109,
                   87.891,
                   1.876,
                   13.9,
                   7.048,
                   0,
                   0,
                  1,
                   0,
                   0,
                   0,
                   0,
                   0,
                   0,
                   0,
                   124.237,
                   123.164,
                   144.439,
                   114.506]).reshape(1,-1)
linha1 = sc_X.transform(linha1)
y_pred_mse_4 = classifier.predict(linha1)
print('Predict Test for Dhaka (2008) : ')
print('Predicted value: ',y_pred_mse_4)
print('Real value: ','1.01')
```

```
Predict Test for Dhaka (2008) :
Predicted value:  [[1.6164486]]
Real value:  1.01
```

Figure 5.22: Prediction test for Dhaka district 2008

This is another prediction test for Dhaka 2008. The result was a bit fluctuated from the actual but still we managed to achieve closer as possible. The predicted value for Dhaka district was about 1.61 and the real value given was 1.61.

5.23 Deep Neural Network system with Root Mean Squared Error

This is the second loss function that we have used for our model. We didn’t changed the density of the neurons in our hidden layer and moreover activation function was set as previous which was ‘tanh’ and kept the validation set 20 percent. Finally we did epoch for 100 times to check how much loss we got.

```
Loss:      0.2678666412830353
Val_loss:   0.4195789396762848
```

Figure 5.23.1:Loss and Validation loss result

```
14/14 [==============================] - 0s 256us/step
Test Score: 0.3531104624271393
56/56 [==============================] - 0s 95us/step
Train Score: 0.2997417577675411
```

Figure 5.23.2: Evaluating model in the train and test score result

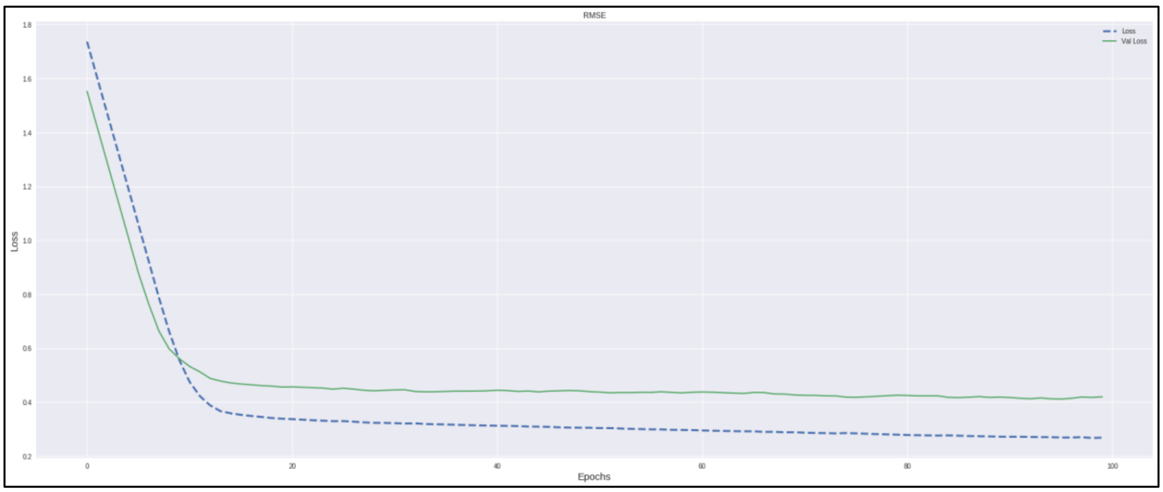

Figure 5.23.3: Loss and Validation loss graph

From the above graph its representing that till 10 epochs the loss was significantly decreasing but after that the loss didn't decreased as expected so after 100 epochs we managed to achieve loss about 0.267 and validation loss about 0.42. This result shows us that there might be a high chance that our data may get over fitted.

5.24 Result Analysis of prediction on Mymenshing district 2010

```
linha1 = np.array([2010,
                   49.829,
                   2.095,
                   30.7,
                   21.3,
                   71.548,
                   475.919,
                   908.157,
                   2.425,
                   36.304,
                   39.786,
                   66.3,
                   119.974,
                   32.347,
                   49.366,
                   29.182,
                   163.675,
                   148.743,
                   68.689,
                   9.442,
                   0,
                   50.861,
                   0,
                   9.667,
                   33.632,
                   0,
                   0,
                   0,
                   90.193,
                   125.976,
                   0,
                   13.458,
                   18.326,
                   30.396,
                   2.69,
                   1.285,
                   10.866,
                   7.307,
                   0.377,
                   17.103,
                   391.055,
                   382.617,
                   321.083,
                   333.083
                  ]).reshape(1,-1)
linha1 = sc_X.transform(linha1)
y_pred_rmse_1 = classifier.predict(linha1)
print('Prediction Test for Mymenshing (2010)' )
print('Predicted value: ',y_pred_rmse_1)
print('Real value: ','1.44')
```

```
Prediction Test for Mymenshing (2010)
Predicted value:  [[1.6330595]]
Real value:  1.44
```

Figure 5.24: Prediction test for Mymenshing district 2010

This figure shows us how much accuracy we got after using the RMSE loss function. We considered genuine 46 Mymenshing district 2010 parameters value as the input and predicted our output. The actual value was about 1.44 and our model predicted 1.63 which seems better than the previous loss function that we have used. But we thought only depending on this result is not sufficient enough to come up to a conclusion. So, we decided to test same thing for other district as well.

5.25 Result Analysis of prediction on Narayangonj district 2015

```
linha1 = np.array([2015,
                   0.021,
                   1.633,
                   28.3,
                   15.2,
                   0.049,
                   12.516,
                   101.838,
                   0.397,
                   62.389,
                   3.777,
                   62,
                   17.964,
                   9.125,
                   7.798,
                   0.711,
                   9.076,
                   13.243,
                   14.834,
                   22.629,
                   0,
                   14.684,
                   0,
                   2.245,
                   2.915,
                   0,
                   0,
                   0,
                   0,
                   21.332,
                   27.424,
                   0,
                   0.504,
                   0,
                   2.314,
                   5.156,
                   0.021,
                   0.192,
                   0,
                   0.683,
                   60.077,
                   58.939,
                   57.837,
                   55.391
                  ]).reshape(1,-1)
linha1 = sc_X.transform(linha1)
y_pred_rmse_2 = classifier.predict(linha1)
print('Prediction Test for Narayangonj (2015)' )
print('Predicted value: ',y_pred_rmse_2)
print('Real value: ','2.352')

Prediction Test for Narayangonj (2015)
Predicted value:  [[1.7406384]]
Real value:  2.352
```

Figure 5.25: Prediction test for Narayangonj district 2015

Next we tried to get see how accurate our model can predict for Narayangonj district 2015. What we have achieved wasn’t actually performed better than RMS loss function. The actual value was 2.352 and predicted value was 1.74.

5.26 Result Analysis of prediction on Dhaka district 2014

```
linha1 = np.array([2014,
                   0.477,
                   1.399,
                   31.1,
                   14.2,
                   0.948,
                   25.676,
                   200.56,
                   0.71,
                   41.062,
                   52.2,
                   64.7,
                   35.321,9.233,
                   5.369,1.926,
                   30.118,32.245,20.539,
                   41.335,0,30.137,4.952,6.832,6.297,
                   7.703,26.322,28.199,87.938,1.896,14.095,
                   7.042,0,0,0,0,1,0,0,0,124.158,124.955,140.524,
                   112.11
                  ]).reshape(1,-1)
linha1 = sc_X.transform(linha1)
y_pred_rmse_3 = classifier.predict(linha1)
print('Prediction Test for Dhaka (2014)' )
print('Predicted value: ',y_pred_rmse_3)
print('Real value: ','1.988')
```

```
Prediction Test for Dhaka (2014)
Predicted value:  [[1.4751384]]
Real value:  1.988
```

Figure 5.26: Prediction test for Dhaka district 2014

From the above figure it can be seen that Dhaka city performed well for this model. Here we are seeing compare to MSE this is performing badly in predicting. The predicted value is about 1.47 whereas the actual value was 1.988.

5.27 Result Analysis of prediction on Dhaka district 2008

```
# Predicting the 15 value at test set
newValues = np.array([2008,
                    0.793,
                    2.385,
                    34.2,
                    12.5,
                    .804,
                    9.691,
                    233.939,
                    1.129,
                    33.679,
                    29.825,
                    8.262,
                    4.808,
                    1.573,
                    30.118,
                    32.245,
                    20.539,
                    41.335,
                    0,
                    30.137,
                    4.869,
                    6.787,
                    6.253,
                    7.691,
                    26.109,
                    28.109,
                    87.891,
                    1.876,
                    13.9,
                    7.048,
                    0,
                    0,
                   1,
                    0,
                    0,
                    0,
                    0,
                    0,
                    0,
                    0,
                    124.237,
                    123.164,
                    144.439,
                    114.506
                      ]).reshape(1, -1)
linha1 = sc_X.transform(linha1)
y_pred_rmse_4 = classifier.predict(linha1)
print('Prediction Test for Dhaka (2008)' )
print('Predicted value: ',y_pred_rmse_4)
print('Real value: ','1.01')
```

```
Prediction Test for Dhaka (2008)
Predicted value:  [[1.3507535]]
Real value:  1.01
```

Figure 5.27: Prediction test for Dhaka district 2008

This figure shows that the predicted value we can see suddenly performed better in RMSE loss function. The predicted value is 1.35 whereas the actual result is 1.01. So to conclude, we can show that MSE actually performing better than RMSE function

5.28.1 Percentage Error of Aman Rice

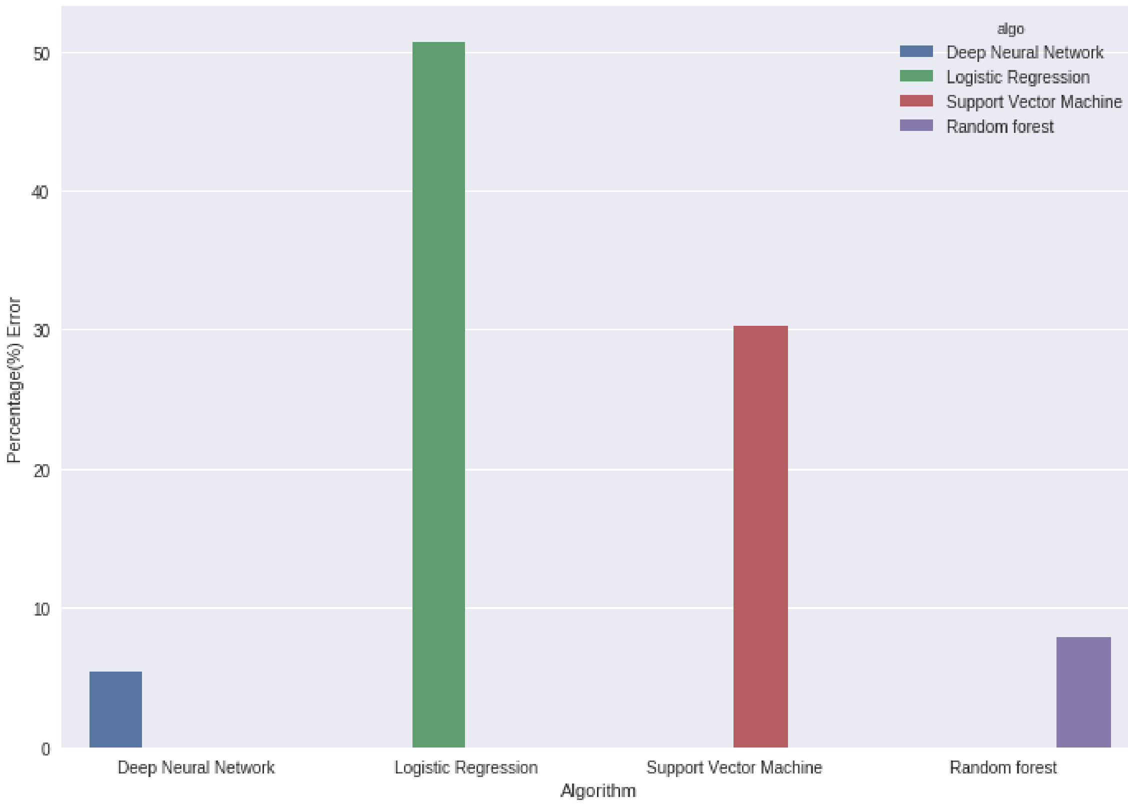

Figure 5.28.1: Percentage Error of Aman Rice for different algorithms

In figure 5.28.1, it is showing that deep neural network is performing much better than the logistic regression and support vector machine. Random forest thus has error under 10 percent.In the horizontal axis, we have set algorithm and in the vertical axis, we have set percentage (%) error. For Aman rice dataset, we have considered 80% training set and 20% as testing set, for all four algorithms. In addition, we got 97.7% accuracy for deep neural network with 2.3% mean square error, 73.3% accuracy for support vector machine with 26.7% mean square error, 90.7% accuracy for random forest with 9.3% mean square error and 52.57% accuracy for logistic regression with 47.43% mean square error. Thus, we are choosing deep neural network approach for crop selection and yield prediction.

5.28.2 Percentage Error of Aus Rice

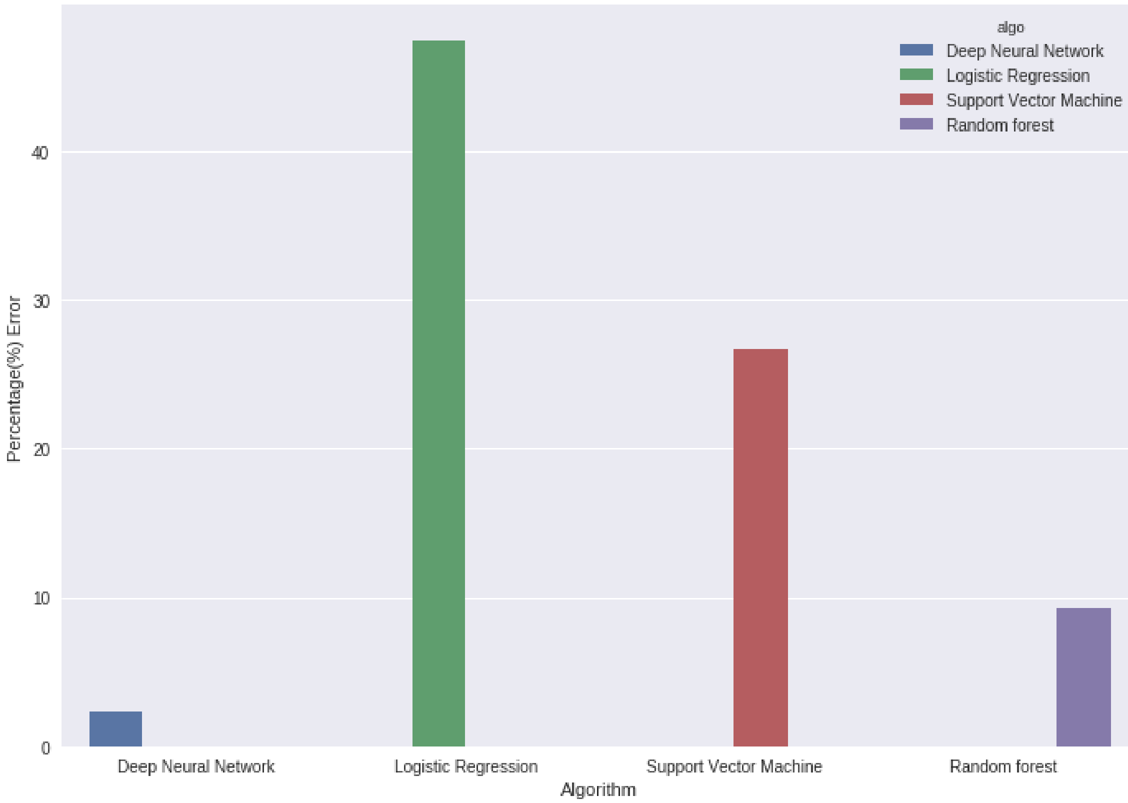

Figure 5.28.2: Percentage Error of AusRice for different algorithms

In figure 5.28.2, it is displaying that deep neural network is performing much better than the logistic regression and support vector machine. Random forest thus has error under 10 percent.In the horizontal axis, we have set algorithm and in the vertical axis, we have set percentage (%) error. For Aus rice dataset, we have considered 80% training set and 20% as testing set, for all four algorithms. In addition, we got 94.6% accuracy for deep neural network with 5.4 % mean square error, 69.7% accuracy for support vector machine with 30.3% mean square error, 92.1% accuracy for random forest with 7.9% mean square error and 49.3% accuracy for logistic regression with 50.7% mean square error. Thus, we are choosing deep neural network approach for crop selection and yield prediction.

5.28.3 Percentage Error of Boro Rice

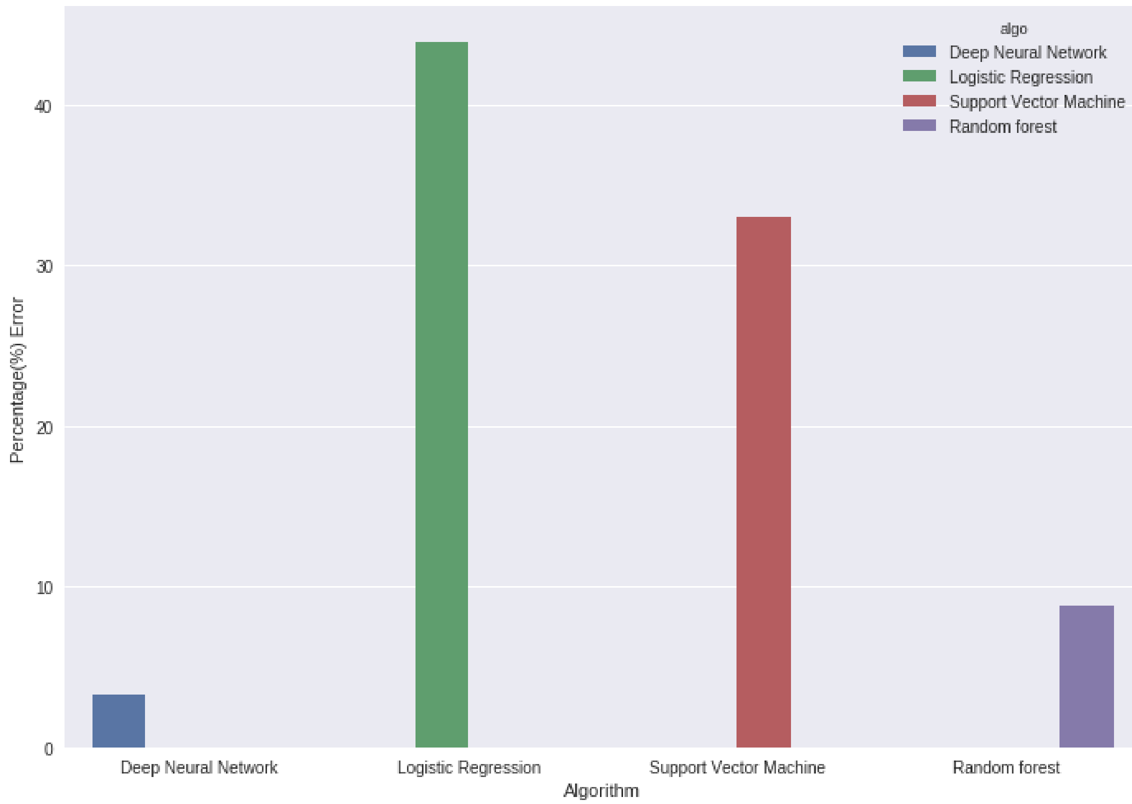

Figure 5.28.3: Percentage Error of Boro Rice for different algorithms

In figure 5.28.3, it is displaying that deep neural network is performing much better than the logistic regression and support vector machine. Random forest thus has error under 10 percent.In the horizontal axis, we have set algorithm and in the vertical axis, we have set percentage (%) error. For Boro rice dataset, we have considered 80% training set and 20% as testing set, for all four algorithms. In addition, we got 96.7% accuracy for deep neural network with 3.3 % mean square error, 67% accuracy for support vector machine with 33% mean square error, 91.2% accuracy for random forest with 8.8% mean square error and 56.11% accuracy for logistic regression with 43.89% mean square error. Thus, we are choosing deep neural network approach for crop selection and yield prediction.

5.28.4 Percentage Error of Jute

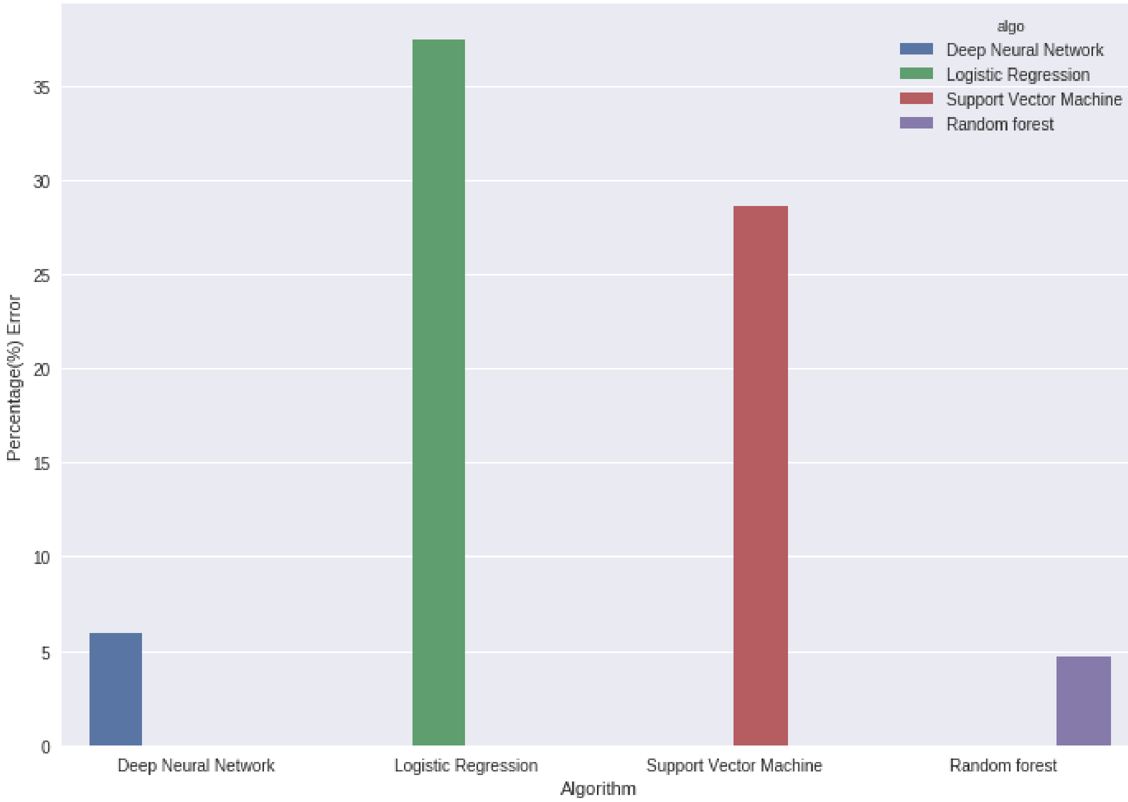

Figure 5.28.4: Percentage Error of Jute for different algorithms

In figure 5.28.4, it is showing that deep neural network is performing much better than the logistic regression and support vector machine. Random forest thus has error under 10 percent. In the horizontal axis, we have set algorithm and in the vertical axis, we have set percentage (%) error. For Jute dataset, we have considered 80% training set and 20% as testing set, for all four algorithms. In addition, we got 94.1% accuracy for deep neural network with 5.9 % mean square error, 71.4% accuracy for support vector machine with 28.63% mean square error, 95.3% accuracy for random forest with 4.7% mean square error and 62.56% accuracy for logistic regression with 37.44% mean square error. Thus, we are choosing deep neural network approach for crop selection and yield prediction.

5.28.5 Percentage Error of Potato

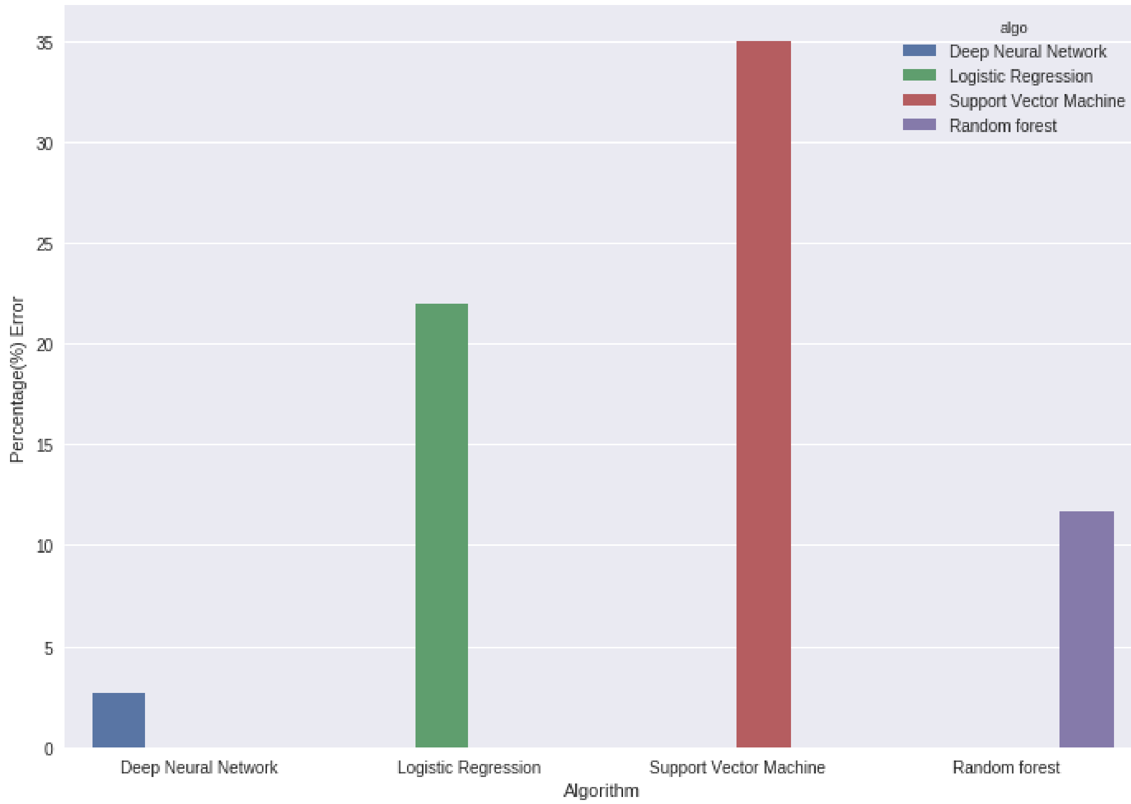

Figure 5.28.5: Percentage Error of Potato for different algorithms

In figure 5.28.5, it is displaying that deep neural network is performing much better than the logistic regression and support vector machine. Random forest thus has error under 10 percent. In the horizontal axis, we have set algorithm and in the vertical axis, we have set percentage (%) error. For Potato dataset, we have considered 80% training set and 20% as testing set, for all four algorithms. In addition, we got 97.3% accuracy for deep neural network with 2.7 % mean square error, 65% accuracy for support vector machine with 35% mean square error, 88.3% accuracy for random forest with 11.7% mean square error and 78% accuracy for logistic regression with 22% mean square error. Thus, we are choosing deep neural network approach for crop selection and yield prediction.

5.28.6Percentage Error of Wheat

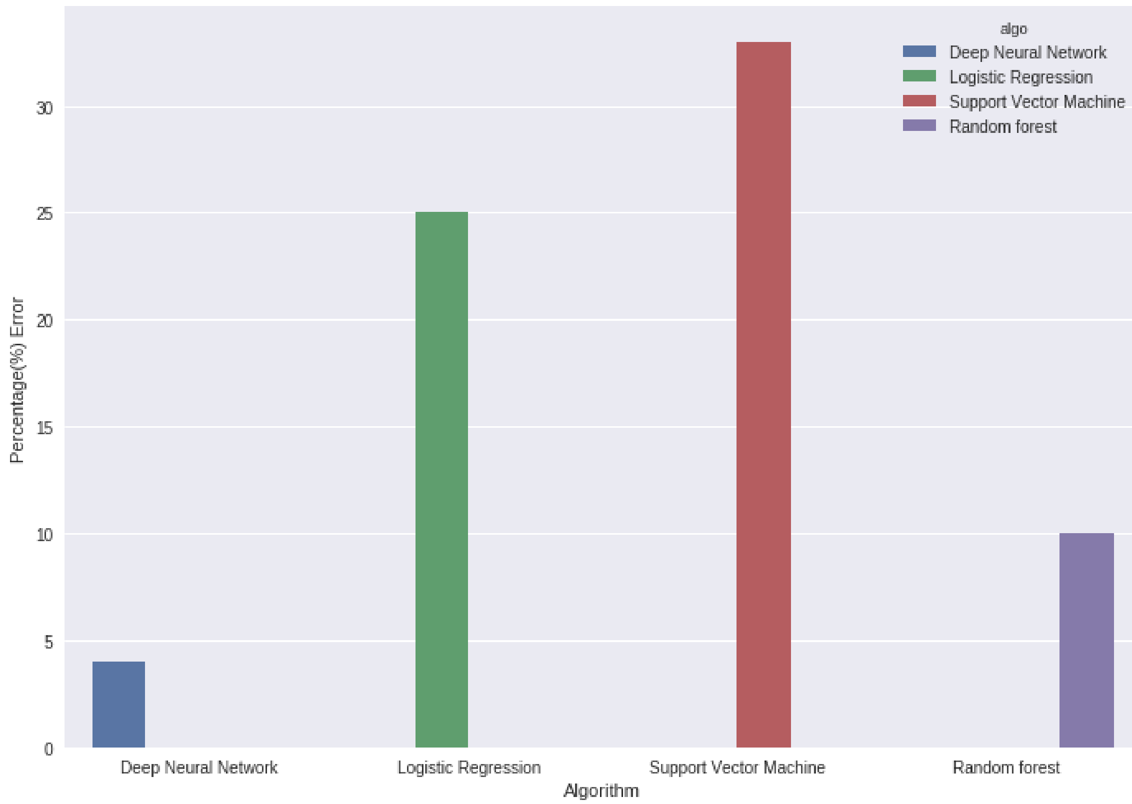

Figure 5.28.6: Percentage Error of Wheat for different algorithms

In figure 5.28.6, it is displaying that deep neural network is performing much better than the logistic regression and support vector machine. Random forest thus has error under 10 percent. In the horizontal axis, we have set algorithm and in the vertical axis, we have set percentage (%) error. For Wheat dataset, we have considered 80% training set and 20% as testing set, for all four algorithms. In addition, we got 96% accuracy for deep neural network with 4% mean square error, 67% accuracy for support vector machine with 67% mean square error, 90% accuracy for random forest with 10% mean square error and 75% accuracy for logistic regression with 25% mean square error. Thus, we are choosing deep neural network approach for crop selection and yield prediction.

5.29Evaluation measures for proposed method

Here we have attached the tabular form of our result analysis.

Table 4:Evaluation measuresof Aus rice

Method	Training(%)	Testing(%)	Accuracy(%)	MSE(%)
Deep Neural Network(DNN)	80%	20%	97.7%	2.3%
Support Vector Machine(SVM)	80%	20%	73.3%	26.7%
Random Forest	80%	20%	90.7%	9.3%
Logistic Regression	80%	20%	52.57%	47.43%

Table 5:Evaluation measuresAman rice

Method	Training(%)	Testing(%)	Accuracy(%)	MSE(%)
Deep Neural Network(DNN)	80%	20%	94.6%	5.4%
Support Vector Machine(SVM)	80%	20%	69.7%	30.3%
Random Forest	80%	20%	92.1%	7.9%
Logistic Regression	80%	20%	49.3%	50.7%

Table 6: Evaluation measuresBoro rice

Method	Training(%)	Testing(%)	Accuracy(%)	MSE(%)
Deep Neural Network(DNN)	80%	20%	96.7%	3.3%
Support Vector Machine(SVM)	80%	20%	67%	33%
Random Forest	80%	20%	91.2%	8.8%
Logistic Regression	80%	20%	56.11%	43.89%

Table 7: Evaluation measures Potato

Method	Training(%)	Testing(%)	Accuracy(%)	MSE(%)
Deep Neural Network(DNN)	80%	20%	97.3%	2.7%
Support Vector Machine(SVM)	80%	20%	65%	35%
Random Forest	80%	20%	88.3%	11.7%
Logistic Regression	80%	20%	78%	22%

Table 8: Evaluation measures Wheat

Method	Training(%)	Testing(%)	Accuracy(%)	MSE(%)
Deep Neural Network(DNN)	80%	20%	96%	4%
Support Vector Machine(SVM)	80%	20%	67%	33%
Random Forest	80%	20%	90%	10%
Logistic Regression	80%	20%	75%	25%

Table 9: Evaluation measures jute

Method	Training(%)	Testing(%)	Accuracy(%)	MSE(%)
Deep Neural Network(DNN)	80%	20%	94.1%	5.9%
Support Vector Machine(SVM)	80%	20%	71.4%	28.6%
Random Forest	80%	20%	95.3%	4.7%
Logistic Regression	80%	20%	62.56%	37.44%

CHAPTER 6

Conclusion and Future work

The purpose of this section is to give readers a sense about implementing the 21^{st} century's modern approach in the agriculture sector and how much scope is available to improve more in this section. Here we are going to look at some of the possible modification that can be done in our research. Moreover, what we have missed, future scope of our research, what more could be added that would help our government and farmers even beyond than we can imagine will be summarized down below.

6.1 Future Works

Many different adaptions, tests and experiments have been left for the future due to lack of time. Future works concern deeper analysis on particular appliance and methods. Here are some of the interesting things that we will like to include in our project later on.

- Soil Moisture Active Passive satellite (SMAP): Our thesis is currently done on supervised data. It would really be a great feature if we can include this remote sensing data, which will continuously send through satellite, which is our first primary goal.
- Robotics: We would like to make a humanoid where that robot will have many important sensors integrated with it for live measurement of any field data. This robot will already have its previous data store in its memory and if any new data is being added the robot will automatically calculate and give prediction with the help of its previous data and existing dataset.
- Data Size: Currently we are working on 0.3 million data set. But this world is not static and everyday billions of data is adding up. The more data we will use for our training the better model we can make to get good accuracy.
- Advanced Very High Resolution Radiometer (AVHRR): This is a broad-band, four or five channel scanner, sensing in the visible, near infrared and thermal

infrared. This data will provide the opportunists for studying and monitoring vegetation conditions in ecosystems.

- Cross-platform: We will like to make cross-platform software, which can be used by our farmers from any device. At the very beginning, we will make a user interface where considered four important fertilizers, which are minerals these are Urea, Triple superphosphate, Diammonium phosphate, and Mp. Moreover, we have categorized thefarmers can choice what type of crops they want to yield based on the current giving input parameters the application will automatically predict the best suitable crop for that particular area.

6.2 Conclusion

To sum up, in our research we tried to develop a model in agricultural sector by using the 21^{st} century's modern approach. Our thesis tried to give a prediction on how simple machine learning algorithm can change our country's agriculture image. Being dependent on agricultural side for a long time our country haven't meet much between agriculture and technology so far. Although there are some mobile apps, which are being, build but which is not actually up to the mark. Right now, our generation people are in a position where everyone is being touched with modern things. So it is high time we should aim at a future to live a better world. Our government has already taken so many good initiatives in agricultural sector. It is high time to precede digitally in this sector so that; not only government but also stockholder and society might get benefitted out from it. Our one little step will be enough to introduce digital agriculture system for best crop selection and yield prediction.

References

[1] Shastry, K., Sanjay, H. &Deshmukh, A. (2016). A Parameter Based Customized Artificial Neural Network Model for Crop Yield Prediction. *Journal of Artificial Intelligence,9*, 23-32 doi: 10.3923/jai.2016.23.32

[2]Patil, D. &Shirdhonkar, M. (2017). Rice Crop Yield Prediction using Data Mining Techniques: An Overview. *International Journal of Advanced Research in Computer Science and Software and Software Engineering, 7*(5), 427-431

[3] Rahman, M., Haq, N. &Rahman, R. (2014). Machine Learning Facilitated Rice Prediction in Bangladesh. *Annual Global Online conference on Information and Computer Technology, 5,*1-4.doi: 10.1109/GOCICT.2014.9

[4] Trading Economics. (2018). Bangladesh GDP Growth Rate.Retrieved from https://tradingeconomics.com/bangladesh/gdp-growth

[5] Trading Economics. (2015). Bangladesh-Agricultural land (% of land area).Retrieved from https://tradingeconomics.com/bangladesh/agricultural-land-percent-of-land-area-wb-data.html

[6] Trading Economics. (2016). Bangladesh-Agricultural value added (annual % growth). Retrieved from https://tradingeconomics.com/bangladesh/agriculture-value-added-annual-percent-growth-wb-data.html

[7] A.I. Wiki. (n.d.).Neural Network.Retrieved fromhttps://skymind.ai/wiki/neural-network

[8] Kotasiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques.*Informatic, 31,* 249-268.

[9] Svozil, D., Kvanička, V., &Pospichal, J. (1997). Introduction to multi-layer feed forward neural networks. *Chemometrics and intelligent laboratory systems*, *39*, 43-62.

[10] Bangladesh Bureau of Statistics. (2008-2017). Bangladesh Bureau of Statistics. Retrieved from http://www.bbs.gov.bd/

[11] Woodford, C. (2018). Neural networks. Retrieved from https://www.explainthatstuff.com from/introduction-to-neural-networks.html

[12] Neilsen, M. (2017). How the back propagation algorithm works. Retrieved fromhttp://neuralnetworksanddeeplearning.com/chap2.html

[13] Sharma, A. (2017). Understanding Activation Function in Neural Networks.Retrieved fromhttps://medium.com/the-theory-of-everything/understanding-activation-functions-in-neural-networks-9491262884e0

[14] Patel, S. (2017) Chapter 2: SVM (Support Vector Machine) – Theory. Retrieved fromhttps://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72

[15] Medcalc. (n.d.).Logistic regression.Retrieved from https://www.medcalc.org/manual/

logistic_regression.php

[16] Donges, N. (Feb 22). The Random Forest Algorithm.Retrieved from https://towardsdata science.com/the-random-forest-algorithm-d457d499ffcd

[17] Neilsen, M. (2017). Deep learning.Retrievedfrom http://neuralnetworksanddeeplearning/

[18] Ji, B., Sun, Y., Yang, S. & Wan, J. (2007). Artificial neural networks for rice yield prediction in mountain regions. *Journal of Agricultural Science, 145,* 249-261.

[19] Dahikar, S., Rode, S. &Deshmukh, P. (2015). An Artificial Neural Network Approach For Agricultural Crop Yield Prediction Based on Various Parameters. *International Journal of Advanced Research in Electronics and Communication Engineering*, *4*(1), 94-98.

[20] Gandhi, N., Petkar, O. & Armstrong, L. (2016). Rice Crop Yield Prediction Using Artificial Neural Networks.*IEEE International Conference on Technological innovations in ICT For Agriculture and Rural Development,* 105-110. doi: 10.1109/TIAR.2016.7801222

[21] Lamba, V. & Dhaka V. (2014). Wheat Yield Prediction Using Artificial Neural Network and Crop Prediciton Techniques (A Survey).*International Journal For Research in Applied Science and Engineering Technology, 2*(IX),330-341.

[22] Jabjone, S. &Jiamrum, C. (2013). Artificial Neural Networks for Predicting the Rice Yield in Phimai District of Thailand.*International Journal of Electrical Energy, 1*(3), 177-181.

[23] Akhand, K., Nizamuddin, M. &Roytan, L. (2018). An Artificial Neural Network-Based Model for Predicting Boro Rice Yield in Bangladesh Using AVHRR-Based Satellite Data.*International Journal of Agriculture and Forestry,8*(1), 16-25. doi: 10.5923/j.ijaf.20180801.04

[24] Jabjone, S. &Wannasang, S. (2014). Decision Support System Using Artificial Neural Network to Predict Rice Production in Phimai District, Thailand. *International Journal of Computer and Electrical Engineering, 6*(2), 162-166. doi: 10.7763/IJCEE.2014.V6.814

[25] Mohan, P. &Patil, K. (2017). Crop Cost Forecasting using Artificial Neural Network with feed forward back propagation method for Mysore Region.*International Journal of Innovative Research in Science, Engineering and Technology, 6*(4), 6133-6141 doi: 10.15680/IJIRSET.2017.0604263

[26] Shastry, K., Sanjay, H. &Deshmukh, A. (2016). A Parameter Based Customized Artificial Neural Network Model for Crop Yield Prediction. *Journal of Artificial Intelligence,9*, 23-32 doi: 10.3923/jai.2016.23.32

[27] Bangladesh Meteorological Department. (2008-2017).Bangladesh Meteorological Department .Retrieved from http://www.bmd.gov.bd/

[28] Soil Resource Development Institute. (2008-2017). Soil Resource Development Institute.Retrieved from http://www.srdi.gov.bd/